

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 15, 2003, 12:41:38 ; Search time 5039 Seconds

(without alignments)
3507.235 Million cell updates/sec

Title: US-10-069-062-7

Perfect score: 2230

Sequence: 1 MSKAFSAPGKAFIAGSYLV.....DIEQTEGVLEKPEPDYIGL 432

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh
-Q/cg2.1/USPTO.spool/US10063062/runat_07082003_161503_4068/app_query.fasta_1.583
-DB=GenEmbl -QMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-OCALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10063062 @CCN 1 1 3508 @runat_07082003_161503_4068 -NCRU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2222	99.6	1299	6	AX087879 Sequence
2	2222	99.6	1763	6	AX087878 Sequence
3	2208	99.0	1299	6	AX240804 Sequence
4	2208	99.0	1299	6	AX488723 Sequence
5	877	39.3	577	6	AX087876 Sequence
6	819	36.7	547	6	AX087874 Sequence
7	809.5	36.3	1356	6	BD170980 Process f
8	809.5	36.3	1356	6	BD171098 Process f
9	809.5	36.3	40397	8	SC9959 S.cerevisia
10	783	35.1	2399	8	YSCERG8
11	610.5	27.4	42947	8	SPAC343
12	475	21.3	2396	6	AX441242 Sequence
13	446.5	20.0	137023	2	AC092736 Trypanoso
14	445.5	20.0	1785	8	AF429385
15	379.5	17.0	110000	2	LMF1CHR15.1
16	287	12.9	1037	11	CNS06337
17	273	12.2	119420	8	AC079041
18	245.5	11.0	300050	1	AP004593
19	231	10.4	728	6	AX441245 Sequence
20	228	10.2	253050	1	AP000984
21	225	10.1	9921	1	AB006889 Sulfolobu
22	217.5	9.8	302050	1	AL935257
23	209.5	9.4	3152	1	AF290093
24	209.5	9.4	300225	1	AE016949 Enterococ
25	193	8.7	3159	1	AF290089 Staphyloc
26	190	8.5	4350	1	LHE279018
27	185.5	8.3	143298	8	AC134240
28	181.5	8.1	1077	6	AX144295
29	181.5	8.1	3169	1	AF290091
30	181.5	8.1	4010	6	AF270256 Staphyloc
31	181.5	8.1	4010	6	AX145574 Sequence
32	181.5	8.1	300242	1	AE016745
33	181	8.1	290150	1	AP004824 Staphyloc
34	177.5	8.0	3074	1	AF290095
35	173	7.8	1074	6	AX620326
36	173	7.8	3145	1	AF290087
37	173	7.8	299050	1	AP003131 Staphyloc
38	173	7.8	343590	1	AP003359 Staphyloc
39	170.5	7.6	200050	1	AL591973
40	170.5	7.6	349980	6	AX638811
41	165.5	7.4	6273	6	AX416851 Sequence
42	165.5	7.4	231450	1	AL596163
43	165.5	7.4	349980	6	AX413018
44	165.5	7.4	349980	6	AX415067
45	162	7.3	3714	1	AF269745 Staphyloc

RESULT 1

ALIGNMENTS

AX087879
 LOCUS AX087879 1299 bp DNA linear PAT 17-MAR-2001
 DEFINITION Sequence 6 from Patent WO0114533.
 ACCESSION AX087879
 VERSION AX087879.1 GI:13396872
 KEYWORDS
 SOURCE Candida albicans
 ORGANISM Candida albicans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE
 1 Rosamond, J.D. and Schnell, N.F.
 TITLE Phosphomevalonate kinase (pmk) gene (erg8) from candida albicans
 JOURNAL Patent: WO 0114533-A 6 01-MAR-2001;
 Astrazeneca AB (SE)

FEATURES
 source 1..1299
 location/Qualifiers
 /organism="Candida albicans"
 /mol_type="genomic DNA"
 /db_xref="taxon:5476"

BASE COUNT 439 a 220 c 273 g 367 t

ORIGIN

Alignment Scores:
 Pred. No.: 2,3e-181 Length: 1299
 Score: 2222.00 Matches: 431
 Percent Similarity: 99.77% Conservative: 0
 Best Local Similarity: 99.77% Mismatches: 1
 Query Match: 99.64% Indels: 0
 DB: 6 Gaps: 0

US-10-069-062-7 (1-432) x AX087879 (1-1299)

QY 1 MetSetLyAlapheserlaProGlyLyValaPheLeuAlaGlyLyLeuValleu 20
 DB 1 ATGTCAAAAGCATTTAGTGACCTGGAAAGCATTTCTGCTGTGATTTGGTCTT 60
 QY 21 GluProIleTyAspAlaTyValaThraAlaLeuSerSerArgMechIsaValaIleThr 40
 DB 61 GAGCCATTATATGATGTTATGTGACGACATTCATCAAGATGATGATGATTAACA 120
 QY 41 ProLyGlyThySerLeuLyGluSerArgIleLySleSerSerProGlnPheAlaasn 60
 DB 121 CCAAAAGAACCGATTGAAAGATCTAGATCAAAATTTCTTCACTCCCAATTTGCAAC 180
 QY 61 GlyGluThrGluTyHisIleSerSerAsnThrGluLyProArgGluValaGlnSerArg 80
 DB 181 GGAAGATGGAAATATCATATCATCAATCAATCAAGAGAGCCAGAGAGTTCACTACGC 240
 QY 81 IleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyIleGlnProThrGlu 100
 DB 241 ATAAATCCATTTTAAAGCACTATATTCATGCTTTAGCTATATTCACCGACGAA 300
 QY 101 AlaPheAspLeuGluIleIleIleTySerAspProGlyTyHisSerGlnIleAspThr 120
 DB 301 GCATTTCATCTGAAATCATCATTTTCTCAGACCTCGATATCATTCACAGAAAGATACT 360
 QY 121 GluThrThySerSerAsnGlyGlyTyThrPheLeuTyHisSerArgAlaIleThr 140
 DB 361 GAAACCAAGACATCTCGAATGAGAAACATTTCTTAAATTCATCTCGTGCATTAC 420
 QY 141 GluValaGlyTyThrGlyLeuGlySerSerAlaGlyLeuValSerValaAlaThrSer 160
 DB 421 GAAAGTGAAAAGCCGATTAAGTTTCATCGCAGATTAAGTCAAGTTGTGCAACAG 480
 QY 161 LeuLeuSerHisPheIleProAsnValaIleSerThrAsnLyAspIleLeuHisAsnVal 180
 DB 481 TTAATATCCATTTTATCCCATGTTATCATAGTAAGAAATTAATTTTGCACAAAGTT 540
 QY 181 AlaGlnIleAlaHisCysTyAlaGlnLyLeuIleGlySerGlyPheAspValaAlaThr 200
 DB 541 GCACAGATTGCACATTGTATATGCCCAAAAAGATGAGATCTGGTTGTATGTTCACACT 600

QY 201 AlaIleTyGlyLeuIleValTyArgArgPheGlnProAlaLeuIleAsnAspValPhe 220
 DB 601 GCATTTATAGCTGATGTTATATAGAACATTTTACCGCAGCTTGTATTAATGACGGTTT 660
 QY 221 GlnValleuGluSerAspProGlyLyPheProThrGluLeuLySleIleGlySer 240
 DB 661 CAGGTTCTGAAAAGTGATCTGATGAGAGTTCCCAAGAGCTTGAAMAAATTTGATGAAGT 720
 QY 241 AsnThrGluGlyTyHisIleGlyValArgCysThrLeuProTyGlyIleTySleLeuMetGly 260
 DB 721 AACTGGAAATTCACACATAGATGATGATATTCATACGAAATCAAGTTATTAATGCGT 780
 QY 261 AspValaGlyGlySerGluThrProLyLeuValSerArgValleuGlnTrpLySle 280
 DB 781 GACGTCAAGGGTGGCTCAGAAAACCCCAATTTGATTCAGGATCTCCAAATGAAAAG 840
 QY 281 GluLyPProGluGluSerSerValaTyArgGlnLeuAsnSerAlaAsnLeuGlnPhe 300
 DB 841 GAAAGCCAGAAAGAAAGCTCTGTGTGTGACCAAGCTTAATAGTCCAAATTTACAGTTT 900
 QY 301 MetLyGluLeuArgGluMetArgGluLyTyAspSerAspProGlyThrTyIleLyS 320
 DB 901 ATGAGGAATTTAGGGAAATGCGTCAAAAATACGACTGACACCGACGACTTATTAAT 960
 QY 321 GluLeuAspHisSerValaGluProLeuThrValaAlaIleLyAsnIleArgLySleu 340
 DB 961 GAGTTATGATCATTTCTGTGAGCTTTGACCTTGCGATTAAGAAACATCAGAAAAGGTTA 1020
 QY 341 GlnAlaLeuThrGlnLySerGlyValaProIleGluProAspValaGlnThrGlnLeu 360
 DB 1021 CAAGCATTAACCAAAAATCAAGAGGTTCCAAATGAACTGAAATGCCAAAGCTGTTG 1080
 QY 361 AspArgCysGlnGluIleProGlyCysValaGlyValaValaProGlyAlaGlyTy 380
 DB 1081 GACGTTGTCAGAGATTTCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
 QY 381 AspAlaIleAlaValleuValleuGluGluGlnValaGlyAsnPheLySleuThrLeu 400
 DB 1141 GATGCAATATGCTGTATTAAGTTTGAAGAAATCAAGTGGGAAATTTTAAGCAGAAACTCTT 1200
 QY 401 GluAsnProAspTyPheHisAsnValaTyTrpValaAspLeuGluGlnThrGly 420
 DB 1201 GAAATCCAGATTAATTTTCAATATGTTTATCTGGTGTGATTTGAAAGCAACAGAAAGT 1260
 QY 421 ValleuGluGlyTyProGlyLyAspTyIleGlyLeu 432
 DB 1261 GTACTTGAAGAAAACCAAGACATATATAGTTTA 1296

RESULT 2
 AX087878 1763 bp DNA linear PAT 17-MAR-2001
 LOCUS AX087878
 DEFINITION Sequence 5 from Patent WO0114533.
 ACCESSION AX087878
 VERSION AX087878.1 GI:13396871
 KEYWORDS
 SOURCE Candida albicans
 ORGANISM Candida albicans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE
 1 Rosamond, J.D. and Schnell, N.F.
 TITLE Phosphomevalonate kinase (pmk) gene (erg8) from candida albicans
 JOURNAL Patent: WO 0114533-A 5 01-MAR-2001;
 Astrazeneca AB (SE)

FEATURES
 source 1..1763
 location/Qualifiers
 /organism="Candida albicans"
 /mol_type="genomic DNA"
 /db_xref="taxon:5476"

BASE COUNT 582 a 300 c 349 g 530 t 2 others

ORIGIN

Alignment Scores:

Pred. No.: 3,44e-181 Length: 1763
 Score: 2222.00 Matches: 431
 Percent Similarity: 99.77% Conservative: 0
 Best Local Similarity: 99.77% Mismatches: 1
 Query Match: 99.64% Indels: 0
 DB: 6 Gaps: 0

US-10-069-062-7 (1-432) x AX087878 (1-1763)

QY 1 MetSerLySAlaPheSerAlaProGluLySAlaPheLeuAlaGlyTyrLeuValLeu 20
 Db 89 ATGTCAAAAGCATTTATGTGACCTGGAAGAAGCATTTCTGTGGTGAATTTGGTCTT 148
 QY 21 GIUProIleTyrAspAlaTyrValThraIleuSerSerArgMetHisAlaValIleThr 40
 Db 149 GAGCCAAATTTAGATGCTTATGATGACAGCATTTGTCAACAGATGATCAGATTATACA 208
 QY 41 ProLySGlyThrSerLeuLySGlySerArgIleLySileSerSerProGlnPheAlaAsn 60
 Db 209 CCAAAAGGAACCAAGTTTGAAGAATCTAGAAATCAAAATTTCTTCAACCCAAATTTGCAAC 268
 QY 61 GlyIuTrpGluTyrHisIleSerSerAsnThrGluLySProArgGluValGlnSerArg 80
 Db 269 GAGAAATGGAAATATCATATCATCAATCAAGAGAGCCCAAGAAATTCAGTCACGC 328
 QY 81 IleAsnProPheLeuGluValaThrIlePheIleValLeuAlaTyrIleGlnProThrGlu 100
 Db 329 ATAAATCCATTTTATAGAGCAACTAATTCATCGTTTATGCTTATATCAACCAACGGA 388
 QY 101 AlaPheAspLeuGluIleIleIleTyrSerAspProGlyTyrHisSerGlnGluAspThr 120
 Db 389 GCATTTGATCTTGAATCATCATCTTATCTCAGACCCGATATCATTCACAAAGATATCT 448
 QY 121 GIUThrLySThrSerSerAsnGlyGluLySThrPheLeuTyrHisSerArgAlaIleThr 140
 Db 449 GAAACCAAGACATCTCGATGAGAGAAACCAATTTCTTACCATTTCTGTCCTCAATACC 508
 QY 141 GIUValGluLySThrGlyLeuGlySerSerAlaGlyLeuValSerValAlaThrSer 160
 Db 509 GAAAGTGAAGAAACCGGATTTAGTTCATGCGAGGATTTAGTTCAGTTGTCACCAAGT 568
 QY 161 LeuLeuSerHisPheIleProAsnValIleSerThrAsnLySAspIleLeuHisAsnVal 180
 Db 569 TTATATATCCATTTATCCCAATGTTATCAGTACGAATAAAGATATTTTGCACAACGTT 628
 QY 181 AlaGlnIleAlaHisCysTyrAlaGlnLySValIleGlySerGlyPheAspValAlaThr 200
 Db 629 GCACAGATTCACATGTTATGCCCCAAAAAGATAGATGCTGGGTTGATGTTGCAACT 688
 QY 201 AlaIleTyrGlyLeuIleValTyrArgArgPheGlnProAlaIleIleAsnAspValPhe 220
 Db 689 GCAATTTATGCTCATTTGTATATAGAAATTTTCAAGCAAGCTTTATATATACCTGTTT 748
 QY 221 GlnValIleuGlnSerAspProGluLySAspPheProThrGluLeuLySLeuIleGlnSer 240
 Db 749 CAGGTTCTAGAAAGATGATCTTGAGAAAGTTCCCAAGAGTTGAAAAAATGATGAAAGT 808
 QY 241 AsnTrpGluGluLySHisGlnArgCysThrLeuProTyrGlyIleLySLeuLeuMetGly 260
 Db 809 AACTGGAAATTCAAACATGAAAGATGATCAATTAACATCGAATCAAGTTATTAATGAGT 868
 QY 261 AspValLySGlyLySerglyThrProLySLeuValSerArgValLeuGlnTrpLyS 280
 Db 869 GACGTCAGAGGTGCTCAGAAACCAACCAATTTGATTCACAGATCTCAATGAGAAAAAG 928
 QY 281 GIULySProGluGlnSerSerValIlyrAspGlnLeuAsnSerAlaAsnLeuGlnPhe 300
 Db 929 GAAAAAGCAAGAAAGCTGTTGTGTATGACCAAGCTTAATATAGCCAAATTTACAGTTT 988
 QY 301 MetLySGluLeuArgGluMetArgGluLySArgGluLySArgAspProGluThrTyrIleLyS 320
 Db 989 ATGAAGAAATTTAGGAAATGCGTGAATAAATTCAGCTCAGACCCAGAGCTTATATTA 1048

QY 321 GIULeuAspHisSerValGluProLeuThrValAlaIleLySAsnIleArgLySGlyLeu 340
 Db 1049 GAGTTAGATCATTTCTGTGAGGCTTTGACTGTGCGATTATACAAATCATGAAAGGGTTA 1108
 QY 341 GlnAlaLeuThrGlnLySerglyValProIleGluProAspValGlnThrGlnLeu 360
 Db 1109 CAAGCATTAACCAAAAAATCAAGAGGTTCAAATGAACTGAAATGTCAAACCAAGTGTG 1168
 QY 361 AspArgCysGlnGluIleProGlyCysValGlyLySValIleValProGlyValaGlyTyr 380
 Db 1169 GACCTGTTCAGAGATTCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1228
 QY 381 AspAlaIleAlaValLeuValLeuGlnGlnValGlyAsnPheLySGlnLySThrLeu 400
 Db 1229 GATGCAATAGCTGTATGTTAGTGGAAATCAAGTGGAAATTTTACCAAGAACTCTT 1288
 QY 401 GIUAsnProAspTyrPheHisAsnValTyrTrpValAspLeuGluGlnThrGluGly 420
 Db 1289 GAAATTCAGATTAATTTTATATATGTTTACTGGGTTGATTTGGAGAGCAACAGAGT 1348
 QY 421 ValLeuGluGluLySProGluAspTyrIleGlyLeu 432
 Db 1349 GTACTTGAAAGAAACCAAGAGCTATATAGGTTTA 1384

RESULT 3
 AX240804 1299 bp DNA linear PAT 26-SEP-2001
 LOCUS AX240804
 DEFINITION Sequence 42 from Patent WO0160975.
 ACCESSION AX240804
 VERSION AX240804.1 GI:15797740
 KEYWORDS
 SOURCE
 ORGANISM
 Candida albicans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE
 1 Roemer,T., Jiang,B., Boone,C. and Bussey,H.
 Gene disruption methodologies for drug target discovery
 JOURNAL Patent: WO 0160975-A 42 23-AUG-2001;
 Eiltra Pharmaceuticals, Inc. (US)

FEATURES
 source
 1..1299
 location/Qualifiers
 /organism="Candida albicans"
 /mol_type="genomic DNA"
 /db_xref="taxon:5476"

BASE COUNT 439 a 221 c 271 g 368 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3,66e-180 Length: 1299
 Score: 2208.00 Matches: 428
 Percent Similarity: 99.54% Conservative: 2
 Best Local Similarity: 99.07% Mismatches: 2
 Query Match: 99.01% Indels: 0
 DB: 6 Gaps: 0

US-10-069-062-7 (1-432) x AX240804 (1-1299)

QY 1 MetSerLySAlaPheSerAlaProGlyLySAlaPheLeuAlaGlyTyrLeuValLeu 20
 Db 1 ATGTCAAAAGCATTTATGTGACCTGGAAGAAGCATTTCTGTGGTGAATTTGGTCTT 60
 QY 21 GIUProIleTyrAspAlaTyrValThraIleuSerSerArgMetHisAlaValIleThr 40
 Db 61 GAGCCAAATTTAGATGCTTATGATGACAGCATTTGTCAACAGATGATCAGATTATACA 120
 QY 41 ProLySGlyThrSerLeuLySGlySerArgIleLySileSerSerProGlnPheAlaAsn 60
 Db 121 CCAAAAGGAACCAAGTTTGAAGAATCTAGAAATCAAAATTTCTTCAACCCAAATTTGCAAC 180
 QY 61 GlyIuTrpGluTyrHisIleSerSerAsnThrGluLySProArgGluValGlnSerArg 80
 Db 181 GAGAAATGGAAATATCATATCATCAATTAATCAAGAAACCAAGAAATTCAGTCAACG 240

```

QY 81 ILeaSPProPheLeuGluAlaThrIlePheIleValIleuAlaTyrIleGlnProThrGlu 100
DB 241 ATAAATCCATTTTAAAGCACTAATTCATCTGTTTACCTATATTCACAGCA 300
QY 101 AlaPheAspLeuGluIleIleIleTyrSerAspProGlyTyrHisSerGlnIuAspThr 120
DB 301 GCATTGATCTTGAAATCATTTATTTACTCGACCTCGATATCATTCACAGAAAGATCT 360
QY 121 GluThrIleThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 140
DB 361 GAAACCAACACATCTCGAATGAGAAAAAATCTTTCTTACCATCTCGTGCATTTACC 420
QY 141 GluValIGluIleThrGlyLeuGlySerSerSerSerSerSerSerSerSerSerSerSer 160
DB 421 GAAAGTGAAGAAAGCCGATTTAGCTTCATCGCAGCATTTAGTCACTGTTGCCACAGT 480
QY 161 LeuLeuSerHisPheIleProAsnValIleSerThrAsnIleAspIleLeuHisAsnVal 180
DB 481 TTATATTCATTTTATCCCAATGTTATCAGTACCAATTAAGATATTTTGCACACGTT 540
QY 181 AlaGlnIleAlaHisCysTyrAlaGlnIleGlySerGlyPheAspValAlaThr 200
DB 541 GCACAGATTGCACATTTGTTATGCCAAAAAAGATGAGATCTGGGTTGATTTGCAACT 600
QY 201 AlaIleTyrGlyLeuIleValTyrArgArgPheGlnProAlaLeuIleAsnAspValPhe 220
DB 601 GCATTTTATGCTGCTGATTTGATATTAAGATTTTACCCAGCTTGTATTAATGACGCTTT 660
QY 221 GlnValIleuGluSerAspProGluIuysPheProThrGluLeuIleIleGluSer 240
DB 661 CAGGTTCTGAAAGATGATCCTGAGAGATCCCAAGATGGAATAATTTGATTTGCAAGT 720
QY 241 AsnTTPGluIuIleHisIleGluArgCysThrIleuProTyrGlyIleIleIleuMetGly 260
DB 721 AACTGGGAATTCAAACATGAAAGATGATACATTCACACGGAATCAAGTATTAATGAGGT 780
QY 261 AspValIleGlyGlySerGluThrProIleIleValSerArgValIleuGlnTTPlyIle 280
DB 781 GACGTCAAGGCTGCTCAGAAACACCCCAATTTGTTACAGACTCTCCAAATGGAAGAG 840
QY 281 GluIuysProGluIuIleSerSerValIleTyrAspGlnLeuAsnSerIleAsnLeuGlnPhe 300
DB 841 GAAAGCCAGAAAGAAAGCTCTGTTGTGTGACCACTTAATGTCCTCAATTTTACAGTTT 900
QY 301 MetIuysGlnLeuArgIleuIleArgIuIuysTyrAspSerAspProGluThrTyrIleIle 320
DB 901 ATGAAGGAATGTAGGGAAATGCTGTGAAATACGACTCAGACCCAGAGACTTATATTAA 960
QY 321 GluLeuAspHisSerValGluProLeuThrValAlaIleIleAsnIleArgIuysGlyLeu 340
DB 961 GAGTTGATCATCTTCTTGAAGCTTTGACCTTTGCACTTGAAGACATCCAGAAAGGCTTA 1020
QY 341 GlnAlaLeuThrGlnIuysSerGluValProIleGluProAspValGlnThrGlnLeu 360
DB 1021 CAAGCACTTAAACCAAAATCAGAGGTTCCAAATGAACTGATGTCCAAACCCAGTTGTG 1080
QY 361 AspArgCysGlnGluIleProGluIuysValIleGlyValValProIleIleGlyIle 380
DB 1081 GACCGTTGTCAGAGATCTCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
QY 381 AspAlaIleAlaValIleuValIleuGluAsnGlnValIleAsnIleIleGlnIleu 400
DB 1141 GATGCAATAGCTCTATTAAGTTGGAATTAATCAAGTGGAAATTTTAAAGCAAGAACTCTT 1200
QY 401 GluAsnProAspTyrPheHisAsnValTyrTTPValAspLeuGluGlnThrGlnIle 420
DB 1201 GAAATCCAGATATTTTCTATATGTTTACTGGGTTGATTTGGAAGCAAGCAAGGT 1260
QY 421 ValIleuGluGluIuysProGluAspTyrIleGlyLeu 432
DB 1261 GTAATTGAAAGAAACCAAGAGCTATATAGGTTTA 1296

```

```

RESULT 4
AX488723 1299 bp DNA linear PAT 16-AUG-2002
LOCUS AX488723
DEFINITION Sequence 6023 from Patent WO02053728.
ACCESSION AX488723
VERSION AX488723.1 GI:22322735
KEYWORDS
SOURCE
ORGANISM
Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mltosporic Saccharomycetales; Candida.
REFERENCE
1
Roemer, T., Jiang, B., Boone, C., Bussey, H. and Ohlsen, K. L.
Gene disruption methodologies for drug target discovery
Patent: WO 02053728-A 6023 11-Jul-2002;
Eliera Pharmaceuticals, Inc. (US)
FEATURES
source
1..1299
Location/Qualifiers
/organism="Candida albicans"
/mol_type="genomic DNA"
/db_xref="taxon:5476"
BASE COUNT 439 a 221 c 271 g 368 t
ORIGIN
Alignment Scores:
Pred. No.: 3,666-180 Length: 1299
Score: 2208.00 Matches: 428
Percent Similarity: 99.54% Conservative: 2
Best Local Similarity: 99.07% Mismatches: 2
Query Match: 99.01% Indels: 0
DB: Gaps: 6
US-10-069-062-7 (1-432) x AX488723 (1-1299)
QY 1 MetSerIleValAlaPheSerAlaProGlyIuysAlaPheLeuAlaGlyIleTyrIleuValIleu 20
DB 1 ATGTCAAAAGCAATTTAGTGACCTTGAAAGATTTCTTGCTGTGGATATTTGGTCTT 60
QY 21 GluProIleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThr 40
DB 61 GAGCCATTTATGAGCTTATATGACGATTTGTATCATCAGATGACATGCACTTATACA 120
QY 41 ProIuysGlyThrSerLeuIuysGluSerArgIleIleIleIleSerSerProGlnPheAlaAsn 60
DB 121 CCAAAAGGAACAGTTTGAAAGATCTGAATCAAAATTTCTTCAACCCCAATTTGCAAC 180
QY 61 GlyIuTTPGluTyrHisIleSerSerSerSerSerSerSerSerSerSerSerSerSerSer 80
DB 181 GGAGATGGGAATATCATCATATCATCAATCAATCAAGAAAAACCAAGAGATTGACGACCC 240
QY 81 ILeaSPProPheLeuGluAlaThrIlePheIleValIleuAlaTyrIleGlnProThrGlu 100
DB 241 ATAAATCCATTTTAAAGCACTAATTCATCTGTTTACCTATATTCACAGCA 300
QY 101 AlaPheAspLeuGluIleIleIleTyrSerAspProGlyTyrHisSerGlnIuAspThr 120
DB 301 GCATTGATCTTGAAATCATTTATTTACTCGACCTCGATATCATTCACAGAAAGATCT 360
QY 121 GluThrIleThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 140
DB 361 GAAACCAACACATCTCGAATGAGAAAAAATCTTTCTTACCATCTCGTGCATTTACC 420
QY 141 GluValIGluIleThrGlyLeuGlySerSerSerSerSerSerSerSerSerSerSerSer 160
DB 421 GAAAGTGAAGAAAGCCGATTTAGCTTCATCGCAGATTTAGTCACTGTTGCCACAGT 480
QY 161 LeuLeuSerHisPheIleProAsnValIleSerThrAsnIleAspIleLeuHisAsnVal 180
DB 481 TTATATTCATTTTATCCCAATGTTATCAGTACCAATTAAGATATTTTGCACACGTT 540
QY 181 AlaGlnIleAlaHisCysTyrAlaGlnIleGlySerGlyPheAspValAlaThr 200
DB 541 GCACAGATTGCACATTTGTTATGCCAAAAAAGATGAGATCTGGGTTGATTTGCAACT 600

```

QY 201 AAlaIleTyGlyLeuIleValIleTyArgArgPheGlnProAlaIleuIleAsnAspValPhe 220
 Db 601 GGAATTATGCTGATGTATATAGAGATTTCACGCGCTTGTATTAATGACGGTTT 660
 QY 221 GlnValLeuGlnSerAspProGluYysPheProThrGlnLeuIleYsleuIleGlnSer 240
 Db 661 CAGGTTCTAGAAAGATGATCTGAGAGAGTCCCAAGAGTTCGAAATTAATGATTCAGAGT 720
 QY 241 AAntTPGluGluYysHisGluArgCysThrLeuProTyGlyIleYsleuLeuMetGly 260
 Db 721 AACTGGGAATTCAAACATGATAAGATTAATTAACACAGCAATCAAGTTATTAATGGGT 780
 QY 261 AspValYsGlyGlySerGluThrProYsLeuValSerArgValLeuGlnIleProYsYs 280
 Db 781 GACGTCAAGGCTGGCTCAGAAACACCAAAATGGTATCAAGATTCACAAATGAAAAAG 840
 QY 281 GlnYsProGluGlnSerSerValValIleYsAspGlnLeuAsnSerAlaAsnLeuGlnPhe 300
 Db 841 GAAAGCCAGAAAGAGCTCTGTTGTATGACAGCTCAATTAATGATGCAATTTACAGTTT 900
 QY 301 MetYsGlnLeuArgGluMetArgGluYysTyraAspSerAspProGluThrTyrlleYs 320
 Db 901 ATGAAGCAATTAAGGAAATGCGTGAATAATACAGCTCAGACCCAGAGACTTAATTAATA 960
 QY 321 GlnLeuAspHisSerValGluProLeuThrValAlaIleYsAsnIleArgYsGlyLeu 340
 Db 961 GAGTTAGATTCATTCCTGTTGAGCTTTGACTGTCGATTAAGACATCGAAAGGTTTA 1020
 QY 341 GlnAlaLeuThrGlnYsSerGlnValProIleGluProAspValGlnThrGlnLeu 360
 Db 1021 CAGGATTAACCAAAATCAAGAGTCCAAATGAACTGATGTCAAACCCAGTTGTTG 1080
 QY 361 AspArgCysGlnGluIleProGlyCysValGlyGlyValValProGlyAlaGlyYsTyrl 380
 Db 1081 GACCGGTGCAAGAGATTCCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
 QY 381 AspaIleAlaValLeuValLeuGlnGlnValGlnValAsnPheYsGlnYsThrLeu 400
 Db 1141 GATGCAATGCTGTTATGTTATGTTGAAATCAAGTGGAAATTTTAAGCAGAAACTCTT 1200
 QY 401 GlnAsnProAspTyrlPheHisAsnValTyrlTPValAspLeuGlnGlnIleThrGlnGly 420
 Db 1201 GAAAAATCCAGATTAATTTTCATATGTTTACTGCGTTGATTGAAAGACAAACAGAGGT 1260
 QY 421 ValLeuGlnGluYysProGluYsTyrlleGlyLeu 432
 Db 1261 GTACTTGAAGAAACCAAGAGACTATATATAGGTTTA 1296
 RESULT 5
 AX087876/c 577 bp DNA linear PAT 17-MAR-2001
 LOCUS AX087876
 DEFINITION Sequence 3 from Patent WO0114533.
 ACCESSION AX087876
 VERSION AX087876.1 GI:13396869
 KEYWORDS
 SOURCE
 ORGANISM
 Candida albicans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.
 REFERENCE
 1 Rosemond, J.D. and Schnell, N.F.
 TITLE Phosphomevalonate kinase (pmk) gene (erg8) from candida albicans
 JOURNAL Patent: WO 0114533-A 3 01-MAR-2001;
 Astrazeneca AB (SE)
 FEATURES
 source
 1..577
 /organism="Candida albicans"
 /mol_type="genomic DNA"
 /db_xref="taxon:5476"
 BASE COUNT 167 a 109 c 108 g 190 t 3 others
 ORIGIN

Alignment Scores:
 Pred. No.: 2,086-66 Length: 577
 Score: 877.00 Matches: 182
 Percent Similarity: 94.82% Conservative: 1
 Best Local Similarity: 94.30% Mismatches: 7
 Query Match: 39.33% Indels: 4
 DB: 6 Gaps: 0
 US-10-069-062-7 (1-432) x AX087876 (1-577)
 QY 61 GlyGluTPGluTyrlHisIleSerSerAsnThrGluYysProArgGluValGlnSerArg 80
 Db 573 GGGGATGGGATATCAATATCATATCATCAATTAACNGA-AAACCCAAAGATNACAGACCC 515
 QY 81 ILeasn-ProPheLeuGlnAlaThrIlePheIle-ValLeuAlaTyrlleGlnProThr 99
 Db 514 ATAAATCCATTTTATAGAGCACTATATCCATCGGTTTATGCTTAATTAATTAACCGACC 455
 QY 100 GlnAlaPheAspLeuGluIleIleIleTySerAspProGlyTyrlHisSerGlnGluAsp 119
 Db 454 GAAACATTTGATCTTGAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 395
 QY 120 ThrGluThrTyrlHisSerSerAsnGlyGluYysThrPheLeuTyrlHisSerArgAlaIle 139
 Db 394 ACTGAACCAAGACATCTCGAATGAGAAAAAACTTTCTTATACCATTCCTCGTCCATT 335
 QY 140 ThrGluValGluYysThrGlyLeuGlySerSerIleGlyLeuValSerValAlaThr 159
 Db 334 ACCGAAGGAAACACCGGATTAAGTTCATCGGAGATTAAGTTCAGTTGTTGCCACA 275
 QY 160 SerLeuLeuSerHisPheIleProAsnValIleSerThrAsnYsAspIleLeuHisAsn 179
 Db 274 AGTTATATATCCATTTATCCCAATGTTTACGTACAGATTAATTAATTAATTAATTAAT 215
 QY 180 ValAlaGlnIleAlaHisCysTyrlAlaGlnYsYsIleGlySerGlyPheAspValAla 199
 Db 214 GTTGCAAGATTTGCAATGTTATGTCGCAAAAAAGATAGATTCGGGTTGATGTGCA 155
 QY 200 ThrAlaIleTyrlGlyLeuIleValIleTyraArgArgPheGlnProAlaIleuIleAsnAspVal 219
 Db 154 ACTGCATTTATGCTGATTTGATATAGAGATTTACGACGCTTGTATTAATGACGCG 95
 QY 220 PheGlnValLeuGlnSerAspProGluYysPheProThrGlnLeuIleYsleuIleGln 239
 Db 94 TTTGCGTTCTAGAAAGATGCTCGAGAAATTTCCCAAGAGTTGAAAAAATTAATGCA 35
 QY 240 SerAsnTPGluGluYysHisGluArgCysThr 250
 Db 34 AGTAACGTGGAATTCAAACATGAAAGATGTACA 2
 RESULT 6
 AX087874 547 bp DNA linear PAT 17-MAR-2001
 LOCUS AX087874
 DEFINITION Sequence 1 from Patent WO0114533.
 ACCESSION AX087874
 VERSION AX087874.1 GI:13396867
 KEYWORDS
 SOURCE
 ORGANISM
 Candida albicans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.
 REFERENCE
 1 Rosemond, J.D. and Schnell, N.F.
 TITLE Phosphomevalonate kinase (pmk) gene (erg8) from candida albicans
 JOURNAL Patent: WO 0114533-A 1 01-MAR-2001;
 Astrazeneca AB (SE)
 FEATURES
 source
 1..547
 /organism="Candida albicans"
 /mol_type="genomic DNA"
 /db_xref="taxon:5476"
 BASE COUNT 184 a 81 c 123 g 159 t
 ORIGIN

Alignment Scores:

Pred. No.: 1,84e-61 Length: 547
 Score: 819.00 Matches: 156
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 36.73% Indels: 0
 DB: 6 Gaps: 0

US-10-069-062-7 (1-432) x AX087874 (1-547)

QY 277 GATTPPLYSGLYSLYSPROGLUSERSERVALTYRASPGLNLEUANSERALA 296
 DB 2 CAATGAAAAAGGAAAAAGCAGAAAGAGCTCTGTGTATGACAGCTTAATAGTGCC 61

QY 297 AATLGLNPHETELYSGLULEAARGLUMETARGLNLVETRYRSPSERAPPROGLN 316
 DB 62 AATTTACAGTTTATGAGAAATGAGAAATCGTAAATAACAGCTCAGACCAAGAG 121

QY 317 ThrTYrTlElySGlULeuAPhISerValGluProLeuThrValAlaIleYsaNIle 336
 DB 122 ACTTAATTTAAAGATTAGTCACTCTGTGAGCCTTTGACCTTGGATTAGAAACATC 181

QY 337 ArgLYeGLYleuGlnAlaLeuThrGlnLYSserGluValProIleGluProaPValGln 356
 DB 182 AGAAAGGGTTACAGAGATTACACAAATAAGAGGTTCCATGTGAACCTGATGCCAA 241

QY 357 ThrGlnLeuLeuAspArgCysGlnGlnIleProGlyCysValGlyValValProGly 376
 DB 242 ACCCAAGTGTGAGCGGTGTCAAGAGATTCGTGTGTGTGGCGGTGTCCAGGT 301

QY 377 AlaGLYGLYrAspAlaIleAlaValLeuValLeuGlnuSngInValGlyAsnPhelys 396
 DB 302 GCGGTGATACATGATCAATAGCTGATTAAGTTTGGAAATCAAGTGGAAATTTTAA 361

QY 397 GlnLYeThreUGlnuSnpProaPTrYPheHISAsnValTYrTTPValAspleUGln 416
 DB 362 CAGAAACCTTGTAAATCCAGATTTATTTTCATTAATGTTTACGCGTGTGTTGAAGAG 421

QY 417 GlnThrGlnGlyValLeuGlnLYeProGluAspTYrTlEGLYleu 432
 DB 422 CAAACAGAGGTGTACTGTAAAGAAAAACAGAACTATATAGGTTTA 469

RESULT 7 BDI70980 1356 bp DNA linear PAT 17-JAN-2003
 LOCUS BDI70980
 DEFINITION Process for producing prenyl alcohol.
 ACCESSION BDI70980
 VERSION BDI70980.1 GI:27876792
 KEYWORDS MO 02053746-A/22.
 SOURCE Saccharomyces cerevisiae (baker's yeast)
 ORGANISM Saccharomyces cerevisiae

REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 AUTHORS Ota, C., Obata, S., Muramatsu, M., Nishi, K. and Totsuka, K.
 TITLE Process for producing prenyl alcohol
 JOURNAL Patent: WO 02053746-A 22 11-JUL-2002;
 TOYOTA MOTOR CORP, CHIKARA OTO, SHUSEI OBARA, MASAYOSHI MURAMATSU,
 KIYOHICO NISHI, KAZUHIKO TOTSUKA

COMMENT OS Saccharomyces cerevisiae (yeast)
 PN MO 02053746-A/22
 PD 11-JUL-2002
 PE 20-DEC-2001 WO 2001P011214
 PR 28-DEC-2000 JP 00P 403067
 PI CHIKARA OTO, SHUSEI OBARA, MASAYOSHI MURAMATSU, KIYOHICO NISHI,
 KAZUHIKO TOTSUKA
 PC C12N15/52, C12P7/04, C12N1/19, C12N1/21
 CC Process for producing prenyl alcohol
 FH Key Location/Qualifiers
 FT 1. .1356
 Location/Qualifiers

FEATURES

source 1. .1356
 /organism="Saccharomyces cerevisiae"
 /mol_type="genomic DNA"
 /db_xref="taxon:4932"
 BASE COUNT 427 a 236 c 304 g 389 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3.95e-60 Length: 1356
 Score: 809.50 Matches: 190
 Percent Similarity: 57.70% Conservative: 76
 Best Local Similarity: 41.21% Mismatches: 146
 Query Match: 36.30% Indels: 49
 DB: 6 Gaps: 14

US-10-069-062-7 (1-432) x BDI70980 (1-1356)

QY 3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyLYrLeuValLeuGluPro 22
 DB 13 AGAGCCTTCAGTCCCGCAGGAAAGCGTTACTAGCTGGTGGATATTAGTTTATAGTACA 72

QY 23 IleTYrAspAlaTYrValThraLeuSerSerArgMetHlaAlaValIleThrProlys 42
 DB 73 AAATATGAAGCATTTGTATGTCGAAATATCGCAAGAAATGCAATGAGCTAGCCATCCTTAC 132

QY 43 GlyThrSerLeuLYeGluSer-----ArgIleYsIleSerSerProGlnPheAla 59
 DB 133 GGT---TCATTTCAGACGGCTGTATGATTAAGTTGAAGTCCGCTGTAAGAAACATTTAA 189

QY 60 AangLYeGLYrGlnLYrThIleSerSerAsnThrGlu--LYeProArgGluValGln 78
 DB 190 GATGGGGAGTGGCTGATACATTAAGTCTTAAGTGGCTTCACTTCTTTCATATAGGC 249

QY 79 SerArgIleAsnProPheLeuGlnAlaThrIlePheIleValLeuAlaTYrIleGlnPro 98
 DB 250 GGATCTAGAACCTTTCATTGAAAAAGTTAGTGGCTACGTAAGTTTATAGCTTTAACT 309

QY 99 Thr-----GlnAlaPheAspleUGlnIleIleIleTYrSerAsp 111
 DB 310 AACATGAGACACTACTGCAATGAATAAGTTTCTGTTATTAATGAT-----ATTTCCTGAT 363

QY 112 ProGlyTYrHISerGlnuSnpThrGluThrLYeThrSerSerAsnGlyGluYthr 131
 DB 364 GATGCTTACATTTCTCAGAGGATAGCTTACCGAA-----CATCGTGGCAACAGAA 417

QY 132 PheLeuTYrHISerArgAlaIleThrGluValGluYsThrGlyLeuGlySerAla 151
 DB 418 TTGAGTTTTCATTCGACAGAAATGAAAGTTCCCAAAACAGGGCTGCGCTCTCGGCA 477

QY 152 GlyLeuValSerValValAlaThrSerLeuLeuSerHISpheIle-----ProAsn 168
 DB 478 GGTTAATGACAGTTTAACTACAGCTTGGGCTCTTTTGTATCGACCTGAAAT 537

QY 169 ValIleSerThrAsnLYeAspleUGlnIleuHISAsnValAlaGlnIleAlaHISCyTYrAla 188
 DB 538 AATGTATGACAAATATATAGAAAGTTTCTATATTTAGCACAAGTGGCTCATTTGCAAGCT 597

QY 189 GlnLYeLYeIleGlySerGlyPheAspValAlaThrAlaIleTYrGlyLeuIleValTYr 208
 DB 598 CAGGGTAAATTTGGAAGCGGTTTATGAGGCGGCGACATATGATCTATTCAGATAT 657

QY 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGlu 228
 DB 658 AGAAGATTCACACCGCATTAATCTCTAATTTGCCAGATAT-----GGAGTGCT 708

QY 229 LysPheProThrGlnLeuLYeLYeLeuIle---GluSerAsnTYrGluGluLYeGlu 247
 DB 709 ACTTACGGCAGATTAAGTGGCATTTGGTATGTAAGAGCATGCAATATTACATTA 768

QY 248 ArgCYsThreLeuProTYrGlyIleLYeLeuMetGlyAspValLYeGlySerGlu 267
 DB 769 ACTAACCATTAATCTTGGGATTAATCTTATGATGAGGATGATTAAGAAATGGTTCAAGA 828

```

QY 268 ThrProLysLeuValSerArgValLeuGlnTrpLysLeuLysProGluGlnSerSer 287
DB 829 ACAGTAAACCTGCTCAGAGAGGTAAGAAATGGTATGATTCGATATGCCAGAAAGCTTG 888
QY 288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
DB 889 AAAATATATACAGAACTCATGATCAATGCAAAATTCAGATTTATGAGAGACTATCTAAACTA 948
QY 305 ArgGluMetArgGlnLysTyrAspSerAspProGluThrTyrLysLeuGlnLeuAspHis 324
DB 949 GATCGCTTACAGAGACTCATGATGCAATTCACAGGATCAATATTGATGATCTTTAGAGAG 1008
QY 325 Ser-----ValGluProLeuThrValAlaLeuLysAsn 335
DB 1009 AATGACTGACTGCTCAAAAGATGCTGTAATCAACAGAGTTAGAGATGCACTGTCACAA 1068
QY 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspVal 355
DB 1069 ATTAGACGTTCTTTAGAAATAAATACTAAAGATCGTCCGATATTCGAACTCCCGTA 1128
QY 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
DB 1129 CAACACTGACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1188
QY 376 GlyAlaGlyGlyTyrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
DB 1189 GGTGCTGCTGCTTATGACCGCCATTCGATGATGATGATGATGATGATGATGATGATGAT 1248
QY 391 GlnValIleGlyAsnPheLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
DB 1249 CAACCGCTCAATGACAAAGA-----TTTTCTAAGGTTCAA 1284
QY 411 TrpValAspLeuGlnGlnGlnThrGlnGlnGlnValLeuGlnGlnLys-----ProGluAspTyr 429
DB 1285 TGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1344
QY 430 Ile 430
DB 1345 CTT 1347

RESULT 8
ED171098 1356 bp DNA linear PAT 17-JAN-2003
LOCUS BDI71098
DEFINITION Process for producing prenyl alcohol.
ACCESSION BDI71098
VERSION BDI71098.1 GI:27876910
KEYWORDS MO 02053747-A/22.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
            Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
            1 (bases 1 to 1356)
REFERENCE
  Otc.C. and Obata,S.
  Process for producing prenyl alcohol
  Patent: WO 02053747-A 22 11-JULY-2002;
  TOYOTA MOTOR CORP,CHIKARA OTO,SHUSEI OBATA
  OS Saccharomyces cerevisiae (yeast)
  PN MO 02053747-A/22
  PD 11-JULY-2002
  PF 20-DEC-2001 WO 2001JP011215
  PR 28-DEC-2000 JP 00P 401701,28-DEC-2000 JP 00P 403067 PR
  18-SEP-2001 JP 01P 282978
  PI CHIKARA OTO,SHUSEI OBATA
  PC G12N15/52,G12P7/04,G12N1/19,G12N1/21
  FH Process for producing prenyl alcohol
  FT Key
  FT source
  FT 1.1356
  FT location/Qualifiers
  FT 1.1356
  FT location/Qualifiers
  FT 1.1356
  FT /organism='Saccharomyces cerevisiae (yeast)'.
  FT /organism='Saccharomyces cerevisiae'
  FT /mol_type='genomic DNA'
  FT /db_xref='taxon:4932'

```

```

BASE COUNT 427 a 236 c 304 g 389 t
ORIGIN
Alignment Scores:
Pred. No.: 3,956-60 Length: 1356
Score: 809.50 Matches: 190
Percent Similarity: 57.70% Conservative: 76
Best Local Similarity: 41.21% Mismatches: 146
Query Match: 36.30% Indels: 49
DB: 6 Gaps: 14

US-10-069-062-7 (1-432) x BDI71098 (1-1356)
QY 3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyTyrLeuValLeuGlnPro 22
DB 13 AGAGCTTCAAGTCCCGCCAGGAAAGCTTACTAGCTGATGATTTAGTTAATACATACA 72
QY 23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys 42
DB 73 AAATATGAGCAATTTGTATGCTGATATTCGCAAGAAATGCAATGCTGTAGCCCATCTTAC 132
QY 43 GlyThrSerLeuLysGlnSer-----ArgIleLysIleSerSerProGlnPheAla 59
DB 133 GGT---TCATTCGAAAGGCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 189
QY 60 AsnGlyGlnTrpGluTyrHisIleSerSerAsnThrGln---LysProArgGluValGln 78
DB 190 GATGGGAGTGGCTGCTACATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 249
QY 79 SerArgIleAsnProPheLeuGlnAlaThrIlePheIleValLeuAlaTyrIleGlnPro 98
DB 250 GGATCTAAGAACCCCTTTCATGAAAGTAATGCTAAGATGATGATGATGATGATGATGATGAT 309
QY 99 Thr-----GluAlaPheAspLeuGlnIleIleIleTyrSerAsp 111
DB 310 AACATGACGACTACTGCAATGAGAAACTGTTGCTGATGATGAT-----ATTTCCTGAT 363
QY 112 ProGlyTyrHisSerGlnGluAspThrGlnLysThrLysSerSerAsnGlyGluLysThr 131
DB 364 GATGCTTCAATTTCTCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 417
QY 132 PheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAla 151
DB 418 TTGAGTTTTCATTTGCAACAGATTTGAAGAGATGATCCCAAAACAGGCTGGCTCCGCGCA 477
QY 152 GlyLeuValSerValAlaIleThrSerLeuLeuSerHisPheIle-----ProAsn 168
DB 478 GGTATGACACAGTTTAACTACAGCTTGGCTCTTTTGTATCGACCTGGAAAT 537
QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
DB 538 AATGATGACAAATATGAGAAAGTATTCATATTTAGACAAAGTTGCTCATTTGTCACACT 597
QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
DB 598 CAGGTAATAAATTTGAAGAGGTTTGAATGATGATGATGATGATGATGATGATGATGATGAT 657
QY 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGlu 228
DB 658 AGAAGATTCACACCGCATATATCTAATTTGCGACATAT-----GGAAGTCT 708
QY 229 LysPheProThrGlnLeuLysLysLeuIle---GluSerAsnTrpGluGlnLysHisGlu 247
DB 709 ACTTACGCGCACTAACTGCGCATTTGTTGATATAGAAAGACTGAATATTAACATTTAA 267
QY 248 ArgCysThrLeuProTyrGlyLysLeuLeuMetGlyAspValLysGlyGlySerGlu 267
DB 769 AGTAACTTATTAACCTTCGCGATTAATTTAGATGATGATGATGATGATGATGATGATGATGAT 828
QY 268 ThrProLysLeuValSerArgValLeuGlnTrpLysGlyLysProGluGlnSerSer 287
DB 829 ACAGTAAACCTGCTCAGAGAGGTAAGAAATGGTATGATTCGATATGCCAGAAAGCTTG 888

```

QY 288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
 Db 889 AAAATATATACAGAACTCCATCATGCAATTTAGATTGATGACATCTTAACAT 948
 QY 305 ArgGluMetArgGlnLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHis 324
 Db 949 GATCGCTTACACGAGCTCATGACGATTACAGGAGCATATATTGAGTCTTGGAGAG 1008
 QY 325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
 Db 1009 AATGACTGATCCTGCTCAAAAGTATCTGAAATATCAGAACTTACAGATGCGATGCCACA 1068
 QY 336 IleArgLysGlnLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspVal 355
 Db 1069 ATTAGACGTTCTTTGAAAAAATACATAAGATCTGTCGCGATATCCAACTCCCGTA 1128
 QY 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
 Db 1129 CAAACCTAGCTTATTGATGATGATTCAGACCTTAAAGAGACTTCTTCTTAATACCT 1188
 QY 376 GlnValGlyValTyrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
 Db 1189 GGTGCTGGTGGTATGACGCGCATTCGATGATTACTACAGACAGATGTTGATTCAGCGCT 1248
 QY 391 GlnValGlyAsnPheLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
 Db 1249 CAAACCGCTATATACAAAAGA-----TTTCTAGAGTTCA 1284
 QY 411 ThrValAspLeuGluGluGlnThrGluGlyValLeuGluGluLys---ProGluAspTyr 429
 Db 1285 TGGCTGATGATGATCTAGCGCTGACTGGGCTGTAGGAAAAAAGATCCGAAACTTAT 1344
 QY 430 Ile 430
 Db 1345 CTT 1347

RESULT 9
 LOCUS SC9959
 DEFINITION S.cerevisiae chromosome XIII cosmid 9959.
 ACCESSION Z49939.1 GI:887599
 VERSION 249939.1 GI:887599
 KEYWORDS dihydrofolate reductase; ERG8; MRE11; MRP44; MTF1; phosphoenolpyruvate kinase; ribonuclease H; ribosomal protein L44; RNH1; transfer RNA-Arg; ubiquitin carboxyl-terminal hydrolase.
 SOURCE Saccharomyces cerevisiae (baker's yeast)
 ORGANISM Saccharomyces cerevisiae
 BUKARYOTA; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 1 (bases 16981 to 22170)
 Venema, J. and Tollervey, D.
 RPS is required for formation of both 18S and 5.8S rRNA in yeast
 EMBL J. 15 (20), 5701-5714 (1996)
 MEDLINE 97051828
 PUBMED 8896463
 REFERENCE 2 (bases 1 to 40397)
 AUTHORS Skelton, J. and Churcher, C.M.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 40397)
 AUTHORS Barrell, B., Rajandream, M.A. and Walsh, S.V.
 TITLE Direct Submission
 JOURNAL Submitted (22-JUN-1995) Saccharomyces cerevisiae chromosome XIII sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge CB10 1RQ E-mail: Barrell@sanger.ac.uk

Notes:
 All CDS over 100 codons have been analysed. CDS that are completely overlapped and those that are overlapped by more than 50% of their length by a larger CDS have been omitted from this analysis.
 Details of the omitted CDS are available on request. The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons and the calculated codon adaptation index (CAI)

FEATURES

source
 is given for each CDS.
 Cosmid 9959 overlapped at 5' by cosmid 8261, embl entry SC8261X, accessionno. Z49809 and at the 3' by cosmid 9408, embl entry SC9408, accession no. Z48756.
 Location/Qualifiers
 1..40397
 /organism="Saccharomyces cerevisiae"
 /mol_type="genomic DNA"
 /strain="A8972"
 /db_xref="taxon:4932"
 /chromosome="XIII"
 /map="13R"
 /clone="cosmid 9959"
 1..3156
 /note="Ym9959.01, len: > 1051, CAI: 0.15, incomplete unknown orf, overlaps and extends cosmid 8261 orf
 /codon_start=1
 /product="unknown"
 /protein_id="CA90190.1"
 /db_xref="GI:887600"
 /db_xref="SWISS-PROT:003661"
 /translation="SSVEIEVEKYSKKLDGSTEKELVPLSTDTTINNSLIGNEDSI
 YSLDDADALISENLTDVPLMEIKTTPKYEIVISEVYSTSEYDNTVAMPQVETSP
 FNMDFPNSLNDYEKKHDLKSTLALAPAPKXDAEVEAGVTSCITSTSGVNTIF
 HTSKETKOVSLDESEYNTFENENTGDNKQSKNPGVANSSTKEDNTDEKYS
 AINYNATGDSGCEDIETASVNEENLRYCEKDMNEAMSSGDECVKONDDSKTOIS
 PSTDPNPOBNDNTERFSTKYKRNLDLDEDESKELTAAEVVDLDEBESDSY
 EDDVADPERGNDGSENNIYKTKKOTGIVPEPEKKNKHEETLEBANVSSVNV
 ONKMDHTVINOEAQANYEAGEKXYIONTPEKHIISIIEIDNALGNMIEERS
 CYEKTHNEAFERRATTTENTKALENNNTNMDQVQACSDSPDDSTAEKVEGSAK
 HNLDIRVSSSEIESEVPLKPSDRENIRSPPIRVGAVKGVADVDESPVKKIDV
 MDESDDNDVDIGTNODIENKNSGTNDAKNSKSSKERDDDEDAVILGVTAAHN
 DGNNSRVINIDPTTNGAYEEDSEVFRQVYQDKENLHSEBPLVDGLSBOQFEKDH
 SENESEFPITIGDITSANIHSNAPDIDRQQLKXNLSLENYSQLIEDSRKQOE
 SDEVNTSEREDLTFEKSNEYKAGAIIEBDESELDISIQHPEBEDLDSNNQERSIR
 ELNSPEBAELYELEIEGPTETAASSKNDDEROGNIPSTDLPSPDKSEYVDSY
 PYSNENITAEKSAPTSPVEYIFSDTNEPMEINDEIPATTEKHKTKNTVTLDD
 RSEHLSNDVNEPNDNSINIKNNGEPEHQAVDIPKVEYKEOEEMPSKSVLEBO
 KPSMELINDKSPERNNDEETREKOTKAKKSKRNKYNRRRRKKTITBGSAAASNT
 KRRRHEKPSRQNTHPSPVDK"

CDS

misc_feature
 1..312
 /note="overlap with cosmid 8261, EMBL SC8261X, Z49809, positions 27890 to 28201"
 3363..4718
 /gene="ERG8"
 3363..4718
 /gene="ERG8"
 /note="Ym9959.02, ERG8 gene, len: 451, CAI: 0.15, SW:ERG8 YEAST P24521, phosphoenolpyruvate kinase; contains PS00627 GHMP kinases putative ATP-binding domain, conflict at C-terminus due to a frameshift with ERG8_YEAST"
 /codon_start=1
 /product="Erg8p"
 /protein_id="CA90191.1"
 /db_xref="GI:887601"
 /db_xref="SWISS-PROT:P24521"
 /translation="MSELRAPSAFGKALAGCYLVLDTRTEAFVVGLSARMAVAHP
 GSLQSDKFEVRVKSQKPRDGBWLTHIPKSGFIPVSIGGSNKPFIETEVIANVFYK
 PMNDYDCNRNLFEVIDIFSDDAVHSGQEDSVTHRGKRRISFHSRLEBPVKAAGSAG
 LVTVLTALASGIVFDLENVDKYREVLHNIQAQVHCOAGKIGSGFVAAAYASIR
 YRRPPALISNLPDGSATGSKLAHLVDEEDMNTTISNLPGLTLMDGDIKXGSE
 TVKLVQKXKMYDSIMPSLKITRELDANSRPNDGSKLDRHETHDDSDQIRESL
 ENNDCTCKRPIETREDAVATIRSRFKITRESGADLEPPVQTSLLDDCOTLKAVLT
 CLIPGAGYDAIVITKODVLDRAQTANDKRFSSKVQWLDVQADGVKKEKDPETLYD
 K"

misc_feature

3810..3842
 /gene="ERG8"
 /note="PS00627 GHMP kinases putative ATP-binding domain"
 complement(4978..6492)
 /note="Ym9959.03c, unknown, len: 504, CAI: 0.15, possible membrane protein"
 /codon_start=1

```
/product="unknown"
/protein_id="CAA90192.1"
/db_xref="GI:887602"
/translation="MTSTRTRRYAQAACACIMCLPFAAGIIIPGPALIKPILISEGVNHE
LDDPDGDLCTAODIKLNTFALSNATNTMALPVGKILDMYGRVGGIIISGLP
LASGNFSAKHLVSLMDPYLVGTTTLTAAGPPVIFSCFOLANSFPQKSTVALALGAS
FISSALFLYRLLYQNPFTLVNSRFTLYLVPLVLAQCLTTPHSSYKTNHIA
KLAEGLDNGRLIEDTSGIIPDEOROSIHALEREDSIIPSPQRKSYLETIVE
DKLQKSGGIFGVHKGSAVEQIKSPFIMLLFALVAMLRINVFATVROEYVLN
DDMLAKNSIFDMLPLGAVSIPRIGILLDTLITLFTSTALIGVGLIPN
SETMNLIGVILVVRPPYTVSDSSVRFGRTRGTGGLSCICGRMSQNLID
KMTHTTFNNPPIINLTVLVFESLTLFTRSQILPKPVENRGLSNVYTI"
complement(6685..7356)
/notes="YM9959.04c, unknown, len: 223, CAI: 0.10, similar
to SW: Dyr_SCHPO P36591, dihydrofolate reductase (29.4%
identity in 218 aa overlap)"
/codon_start=1
/product="unknown"
/protein_id="CAA90193.1"
/db_xref="GI:887603"
/translation="MTKNTLMHGLAQSDYFASKTKGRAMEKIGYLYTPAPNE
FPADYDFLGVADADGENTGVLALENDPSTGGYFTIQTITIDYLANVLENGP
FAGIVGFSQAGVAGVATDFNGLGLTPEPPEFMAVSGFRPOQOYEOYDLH
PISVPSIHVGELDITTEBPAKVGGLYNSCTEDSRILMHSGHFVNSRGRFVKAQW
LQOLT"
/notes="9177
7762..9177
/notes="YM9959.05, unknown, len: 471, CAI: 0.17, similar to
putative ubiquitin carboxyl-terminal hydrolases, eg
SW: Ubq_HUMAN, P40818, probable ubiquitin
carboxyl-terminal hydrolase (31.5% identity in 352 aa
overlap); contains P800973 Ubiquitin carboxyl-terminal
hydrolases family 2 signature 2"
/codon_start=1
/product="unknown"
/protein_id="CAA90194.1"
/db_xref="GI:887604"
/translation="MSICPHIOVONERHSKQVLTGCAARYIINHSAVPEKKEPLMT
KGTCHHEINSGATFWCLOCGFCGCMNHSPFLSHSKOIGIRPINSNGILFPKCEDY
IGNDLINDALIATKYWDVCTKTWPSERRDGLSGILMSTGCMSSILQCLINPY
FLRHSMSQIHNHNCVRSPPDKFCSCALDIYVELGALNTQKASSSTSTNQCTPIY
LITCAMKINQNLAGSQODAPHEWQFIINQIHQSYVLDLPNAKEVSRANKQCECTIV
TYFEGSLSESIYVCGCONNSKTTIDPELDLSIDDKKKVCLDPSFKKEDLKQENY
HGCNSTDQDAIKQIGHKLPVAVLQKRPHILANGSWRKLDDEIPETPYNMCRVC
STREKDKHSENKVPDIITIELGIYSHKGTVAEGHIAFCIKSGQWFFKNDMSVSI
SOEVLKEDQAYLFFYIRQVN"
/notes="P800973 Ubiquitin carboxyl-terminal hydrolases
family 2 signature 2"
complement(9622..11700)
/genes="MRE11"
/complement(9622..11700)
/notes="MRE11"
/notes="YM9959.06c, MRE11 gene, len: 692, CAI: 0.14,
SW: MRE11_YEAST, P32829, required for repair and meiotic
recombination, conflict, this sequence is longer due to a
frameshift with SW: MRE11_YEAST"
/codon_start=1
/product="Mre11p"
/protein_id="CAA90195.1"
/db_xref="GI:887605"
/translation="MDYDPDPTIRILITTDNHYGNENDPTIGDSSKTFPHFVMLAK
KNNVAVVQSGDLFHNKPSKSLYQVLTFLRCMGDPCCELELSDPSQVHFYDEF
TNVAVEDPENTISIVFGISGNDASGSLCPMDIILATLIMHFGVILSVDXIKY
VILFOKSTKIALYGLAAVNDERLFRFKDGVTFEYVMEGEMFNLCHXHTG
HNTAFLEPOFLPDLMDVIMGHEHCINLVHNPINKNDVOPGSSVATSIQEAEO
PKVFLIDIKYGEAPRMPIPLETIRBTPEKSIISQDVPHLPHKDKATSKYLIOAE
EMIRANETTKOKLADDSGDMVAELPKRLILRDYKASPSGTQSPITDYENPRFS
NFGVAVGANNVOPFYKRSFVTSKSKSINGTISIDSDVERKLSESGGELEVTIV
NDLNRKQSLSTPEVLNEAVKRFVDKDEKTKALKEFIISHEISNEVGILSTNEEFRTD
```

```
DAEEMKALIKQYKRVANSVRPPPKENDETNPFAENGCLDSFRSSNREVRTGSDITQS
HVDNSENRIITHISQAESKPTSKRVRTATKKRIAPFSDSTVISAENELGNNDAOD
DVIDENENDIIMWSTDEEDASVGLNGRKTKTTPRPAASFKTASRRGRASRTPTDI
IGSLAKKRR"
complement(join(12007..12245,12393..12450))
/genes="MRPL44"
/complement(join(12007..12245,12393..12450))
/genes="MRPL44"
/notes="YM9959.07c, MRPL44 gene, (YMR44), len: 98, CAI:
0.16, mitochondrial 60S ribosomal protein L44 (YML44)"
/codon_start=1
/product="Mrp14p"
/protein_id="CAA90196.1"
/db_xref="GI:887606"
/translation="MTTFYFQVIVRFPFPGKEAVARVLAATPPQRMNCTQIOSE
IISDYNKPKPLKVTYTKKKEKNEVDPSNMFOELANHFRDRSKQDLKHELMH"
complement(12246..12248)
/genes="MRPL44"
misc_feature
Alignment Scores:
Pred. No.: 3,366-58 Length: 40397
Score: 809.50 Matches: 190
Percent Similarity: 57.70% Conservative: 76
Best Local Similarity: 41.21% Mismatches: 146
Query Match: 36.30% Gaps: 49
DB: 8 Gaps: 14
US-10-069-062-7 (1-432) x SC9959 (1-40397)
QY 3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyGlyTyrLeuValLeuGluPro 22
...:::
DB 3375 AGAGCTTCAGAGCCCGCAGGAAACGTTACTAGCTGGAGATATTGTTTGAATACA 3434
QY 23 IleTyrAspAlaTyrValThrAlaLeuSerSerAlaGlyMetHisAlaValIleThrProLys 42
...:::
DB 3435 AATATTTGAAGCATTTGTATGTCGATTCGCGAATAAGCAATGCAATGCGCCATCTTAC 3494
QY 43 GlyThrSerLeuLysGluSer-----ArgIleLysIleSerSerProGlnPheAla 59
...:::
DB 3495 GGT---TCATTCGCAAGGCTCGATAGTGAAGCGCGTGAAAGTAACAAATTTAA 3551
QY 60 AsnGlyIuTrpGluTyrHisIleSerSerAsnThrIu---LysProArgGluValAlaGln 78
...:::
DB 3552 GATGGGAGTGCGCTGACATATAAGCTTAAAGTGCCTTATTCGATTGCAATAGCC 3611
QY 79 SerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnPro 98
...:::
DB 3612 GGATCTAAGAACCCCTTTCATTGAAAAGTTATCGCTAAGTATTGACTTTAAACCT 3671
QY 99 Thr-----GluAlaPheAspLeuGluIleIleIleTyrSerSer 111
...:::
DB 3672 AACATGACGACTACTGCAATAGAAAACCTTGTCGTTATGAT-----ATTTCCTGAT 3725
QY 112 ProGlyTyrHisSerGlnGluAspThrGluThrIlyrThrSerSerAsnGlyValuLysThr 131
...:::
DB 3726 GATGCTTACCATTCCTCAGAGAGATAGCTTACCGA-----CATCGTGGCAACGAAGA 3779
QY 132 PheLeuTyrHisSerArgAlaIleThrGluValGluTyrThrGlyLeuGlySerSerAla 151
...:::
DB 3780 TTGAGTTTCATTTCGACAGAAATTGAAGATGCCCAAAACAGGGCTGGGCTCCGCGCA 3839
QY 152 GlyLeuValSerValValAlaThrSerLeuLeuSerHisPheIle-----ProAsn 168
...:::
DB 3840 GGTTCAGTCAAGCTTTTACTACAGCTTGCGCTCTTTTGTATCGAGCTGGAATAT 3899
QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
...:::
DB 3900 AATGATAGCAAAATATAGAGAGTATTATTAATTTAGACACAAGTTGCTATGCTAAGCT 3959
QY 189 GlnLysValIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
...:::
DB 3960 CAGGCTAAATTTGGAAGCGGTTGATGTATGACGGCGGACGATATGATATATATATAT 4019
```


QY 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGlu 228
 Db 4020 AGAAGATTCCACCCGATTATCTTAATTCGCCAGATTT-----GGAAGTGC 4070
 QY 229 LysPheProThrGluLeuLysLysLeuIle--GluSerAsnTrpGluLysHisGlu 247
 Db 4071 ACTTACGGCAGTAACCTGGCGCATTTGGTTGATGAAAGAACATGATATTACGATTAA 4130
 QY 248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyValAspValLysGlySerGlu 267
 Db 4131 AGTACCATTACTTCGGATTAATCTTAATGATGAGCGCATATTAAGAAATGGTTCAAGA 4190
 QY 268 ThrProLysLeuValSerArgValLeuGlnTrpLysLysGluLysProGluGluSer 287
 Db 4191 ACGGTAAACTGCTCCAGAAAGTAAATAATGATGATTCGCAATGCCAGAAAGCTTC 4250
 QY 288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
 Db 4251 AAATATATACAGACTCGCATCGCAAAATTCAGATTATGATGATGATGATGATGATGAT 4310
 QY 305 ArgGluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHis 324
 Db 4311 GATCGCTTACACGAGACTCATGACGATTACAGCGATCATATTTGATGATGATGATGATGAT 4370
 QY 325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
 Db 4371 AATGACTGATCCTGCAAAAGTATCCTGAATATCAGAAATGAGATGATGATGATGATGAT 4430
 QY 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspVal 355
 Db 4431 ATTAGACGTTCTCTTGAATAAATACTTAAGAAATTCGTCGCGATTCGAACTCCCGTA 4490
 QY 356 GlnThrGluLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
 Db 4491 CAACACTGACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4550
 QY 376 GlyAlaGlyValTyrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
 Db 4551 GGGCTGGTGGTATATGACCGCATTTGCGATGATTAACGAAAGATGATGATGATGATGATGAT 4610
 QY 391 GlnValGlyAsnLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 410
 Db 4611 CAACCGCTATGACAAAGAA-----TTTCTTAAGATTCA 4646
 QY 411 TrpValAspLeuGluGlnGlnThrGluGlyValLeuGluGluLys---ProGluAspTyr 429
 Db 4647 TGGCTGATGTAACTCAGGCTGACTGGGCTGTTAGAGAAAGAAAGATCCGAAACTTAT 4706
 QY 430 Ile 430
 Db 4707 CTT 4709
 RESULT 10
 YSCERG8 2399 bp DNA linear PLN 27-APR-1993
 LOCUS S.cerevisiae phosphomevalonate kinase gene, complete cds.
 DEFINITION M63648.1 GI:553127
 ACCESSION M63648.1 GI:553127
 VERSION M63648.1 GI:553127
 KEYWORDS phosphomevalonate kinase.
 SOURCE Saccharomyces cerevisiae (baker's yeast)
 ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 REFERENCE 1 (bases 1 to 2399)
 AUTHORS Tsay, Y.H. and Robinson, G.W.
 TITLE Cloning and characterization of ERG8, an essential gene of
 Saccharomyces cerevisiae that encodes phosphomevalonate kinase
 Mol. Cell. Biol. 11 (2), 620-631 (1991)
 JOURNAL 91117228
 MEDLINE 1846667
 PUBMED
 COMMENT On Oct 3, 1994 this sequence version replaced gi:171478.
 Original source text: S.cerevisiae DNA.

FEATURES
 source Location/Qualifiers
 1..2399
 /organism="Saccharomyces cerevisiae"
 /mol_type="genomic DNA"
 /db_xref="taxon:4932"
 gene 767..2121
 /gene="ERG8"
 TATA_signal 767..774
 /gene="ERG8"
 CDS 847..2121
 /gene="ERG8"
 /EC_number="2.7.4.21"
 /codon_start=1
 /product="phosphomevalonate kinase"
 /protein_id="AAA34596.1"
 /db_xref="GI:171479"
 /translation="MSELRARSAPEKALLAGGYLVDTKRYAFVGLSARHAAVAPY
 GSLQSDYFEVAVKSKPFDGEWLVHISPKGFIPLVSGSKNPIEVAVSVYFK
 PMDDYCNRLNLFVIDIFSDDAVHSDQEDVTEHGRNRSLFHSRIEVPKGLSGAG
 LVTVLTALASPFVSDLENNVDKYREVIHNLQVHCOAGKIGSPVVAARVGSIR
 YRRFPALISNLPDIGSATYGSKLALHLEDEDMNTTISNHLPSGLTLMGDIKGSE
 TYKVQVKRMVDSHMPESLKYTELDELHANSRPMGLSKDLRIHETHDDYDSQIFESL
 ERNDCTCKRTEPTEVRDAVATIRSFRTKTESGADIEPPVQTSLLDDCQTLKGVLT
 CLIPGAGGDAIVAVITKQDVDLRAQPLMTKDF"
 BASE COUNT 840 a 415 c 514 g 630 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.57e-57 Length: 2399
 Score: 783.00 Matches: 186
 Percent Similarity: 58.33% Conservative: 80
 Best Local Similarity: 40.79% Mismatches: 150
 Query Match: 35.11% Indels: 40
 DB: 8 Gaps: 14
 US-10-069-062-7 (1-432) x YSCERG8 (1-2399)
 QY 3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyLysTyrLeuValLeuGluPro 22
 Db 859 AAGAGCTTCAAGGCCCGCAAGAAACGTTACTAGCGGTGATATTGATTAGATACA 918
 QY 23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys 42
 Db 919 AAATATGAAGCATTTGTAGTCGATTAATCGCAAGAAATGACATGCTGAGCCCATCTTAC 978
 QY 43 GlyThrSerLeuLysGluSer-----ArgIleLysIleSerSerProGlnPheAla 59
 Db 979 GGT---TCATTCGAAAGGCTCATATGATTGAAGTGGCGGTCAAAATCAATTTAA 1035
 QY 60 AsnGlyGluTrpGluLysIleSerSerAsnThrGlu---LysProArgGluValGln 78
 Db 1036 GATGGGAGTGGCTGTACATTAATGCTCTTAAGAGCTTCATTCCTTCGATAGGC 1095
 QY 79 SerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnPro 98
 Db 1096 GGATCTTAAGAACCTTTTATTAATAAAGTATCGCTAACGATATTGACTTTAACT 1155
 QY 99 Thr-----GluAlaPheAspLeuGluIleIleLeuTyrSerAsp 111
 Db 1156 AACATGAGCACTACTGCATAGAAACTTGTGCTTATTTGAT-----ATTTCCTGAT 1209
 QY 112 ProGlyTyrHisSerGlnGluAspThrLysThrLysSerSerAsnGlyValLysThr 131
 Db 1210 GATGCTTACATTCAGAGAGATAGCGTACGAA-----CATCGTGGCAACGAGA 1263
 QY 132 PheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAla 151
 Db 1264 TTGAGTTTCATTCGACAGATTTGAAGAGTTCCCAAAACAGGCGTGGCTCTCGGCA 1323
 QY 152 GlyLeuValSerValValAlaThrSerLeuLeuSerHisPheIle-----ProAsn 168
 Db 1324 GATTATGACACAGTTTACTACAGCTTGGCTCTTTTGTATCGGACCTGGAAT 1383


```

Qy 169 ValIleSerThrAsnLysAspIleLeuH1a8enValAlaGlnIleAlaH1scYTyra1a 188
Db 1384 AATGTAGCAAAATATAGAGAAAGTATTCTAATTATTCAGCAAGATTCTCATTTGCTCAAGCT 1443
Qy 189 GlnTylValIleGlySerGlyPheAspValAlaThraIleLeuGlyLeuIleValTyr 208
Db 1444 CAGGGTAAATTTGAAAGCGGGTTGATGAGCCGCCGCGATGATGATCTATTCAGATAT 1503
Qy 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGlnLysSerAspProGlu 228
Db 1504 AGAAGATTCACACCCCATTAATCTCTAATTTCGCCAGATAT-----GGAAGTCT 1554
Qy 229 LysPheProThrGlnLeuLysLysLeuIle---GlnSerAsnTrpGlnLysIleGlu 247
Db 1555 ACTTACGCGAGTAATCGCGCATTTGGTTGATGAAGAAAGACTCGAATTTACGATTAA 1614
Qy 248 ArgGlyThrLeuProGlyTyrGlyIleLysLeuLeuMetGlyAspValLysGlySerGlu 267
Db 1615 AGTAACCATTTACCTTCGGGATTTAACTTTATGATGGGCGATATTAAGATGGTTCAAGA 1674
Qy 268 ThrProLysLeuValSerArgValLeuGlnTrpLysGlyLysProGlnLysSerSer 287
Db 1675 AAGAGTAAACTGGTCCAGAAAGTAAATAATGGTATGATTCGATATGCCAAGAACCTTG 1734
Qy 288 ValValTyrAspGlnLeuLysSerAlaAsnLeuGlnPheMet-----LysGlnLeu 304
Db 1735 AAAATATATACAGAACTCGATCGAATTTCTAGATTTATGATGATGATGATGATGATGAT 1794
Qy 305 ArgGlnMetArgGlnLysTyrAspSerAspProGlnTrpTyrIleLysGlnLeuAspHis 324
Db 1795 GATCGCTTACAGAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1854
Qy 325 Ser-----ValGlnProLeuThrValAlaIleLysAsn 335
Db 1855 AATGACTGACTGCTGCAAAAGTATCTGAAATCAAGAAATTAGAGATCAGATGCTGCCACA 1914
Qy 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGlnValProIleGlnProAspVal 355
Db 1915 ATTAGAGCTTCTCTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1974
Qy 356 GlnThrGlnLeuLeuAspArgCysGlnGlnIleProGlyCysValGlyValValPro 375
Db 1975 CAACATGACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2034
Qy 376 GlyAlaGlyLysTyrAspAlaIleAlaValLeuValLeuGlnAsnGlnValLysPhe 395
Db 2035 GGTGCTGGTGGTATGAGCCCATTTGCAATGATTACT-----AAGCAAGATGTTACTT 2088
Qy 396 LysGlnLysThrLeuGlnLysAspPheTyrPheH1a8enValTyrTrpValAspLeuGlu 415
Db 2089 AAGGGCTCAACCGCTATGACAAAGATTTCTAAGGTTCAATG-----GCTGATGTAAT 2142
Qy 416 GlnGlnThrGlnGlyValLeuGlnGlnLys---ProGlnAspTyrIle 430
Db 2143 CAGGCTGACTGGGTGTTAGAAAGAAAGATTCGGAATACTTACTT 2190

```

RESULT 11
SPAC343 42947 bp DNA linear PIN 23-MAR-2001
LOCUS S.pombe chromosome I cosmid c343.
DEFINITION AL109739
ACCESSION AL109739
VERSION 1
KEYWORDS Aconitase family; CHC4 type (RING finger); G-beta repeat; glutamyl-tryptophan transferase; homoaconitase hydratase (EC 4.2.1.36); methylenetetrahydrofolate reductase 2; mitochondrial ribosomal protein; mchfr2; PRT12 family; phosphomevalonate kinase; rds1; RNA-binding protein; stress response protein lrd1p; transcriptional regulator; translation initiation factor eif-2b beta subunit; tRNA Gly anticodon GGC; tRNA isopentenyltransferase; trp-asp repeat protein; UBX domain; vacuolar atp synthase catalytic subunit a; yma1; WD domain; W40; yeast shp1 homolog; Zinc finger. Schizosaccharomyces pombe (fission yeast)

SOURCE ORGANISM Schizosaccharomyces pombe

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 42947)
Murphy, L., Harris, D., Wood, V., Rajandream, M.A. and Barrell, B.G.
Direct Submission
Submitted (03-AUG-1997) European Schizosaccharomyces genome
sequencing project, Sanger Centre, The Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
Notes:
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
(URL, <http://www.sanger.ac.uk/projects/S.pombe/>)
During 1995 to 1996 about 66% of S. pombe chromosome 1 was
sequenced by the Sanger Centre. The sequencing from The European
genome is now being continued with funding from The European
Commission. Fourteen European sequencing laboratories, including
the Sanger Centre, are participating in the project.
protein coding regions (CDS) have been predicted with the help of
computer analysis using the GeneFINDER program in Pombase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Splice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
number of introns/exons or we may not have chosen the correct
splice donor/acceptor sites. CDS are numbered using the following
system eg SPBC25H2.01c. SP (S. pombe), B (chromosome 2), c25H2
(cosmid name), .01 (first CDS), c (complementary strand)
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid c343
overlaps pl clone p27G11 at the 5' end and cosmid clone c824 at the
3' end.

FEATURES

source

misc_feature
1..2347
/note="nominal overlap with plasmid SPAP27G11, EM:AL157917
S. pombe chromosome 1"
/complement(join(61..1333,1400..1407))
/gene="SPAC343.01c"
complement(join(61..1333,1400..1407))
/gene="SPAC343.01c", len:427, SIMILARITY: Saccharomyces
cerevisiae, ERG8_YEAST, phosphomevalonate kinase, (451
aa), fasta scores: Opt: 407, E(-):9.9e-20, (32.6% identity
in 451 aa)
/codon_start=1
/label="SPAC343.01c
/product="putative phosphomevalonate kinase"
/protein_id="CAB52264.1"
/db_xref="GI:5706502"
/db_xref="SPRMBL:O9UT88"
/translation="MKYTCGARGGVLLAGGVIVLDPQSGIVLGTAKGASTTLD
KCGIVRKSPFPAEMALNIDTVTSIRHQAIEKELRNPPFQALFYINIF
AMTISRLQWODLVTLQVDNAYHQPDLKPDQTSYKRFNLNCTLQVHTGAGSSA
FATISLGLFLSRLRLTDGDSLKTIDSTKYVNLQIARCSAQKSGSPDGA
AAATGSGTIRFDPKLEQLVLPDDEIKINIFTELRKQIVSKMSDVPOLPATYC
LAMDVGSGSTPEPMVKVOOWKQWESKNCPPDLYSRVLSIKNQLSESDSL
OSGSRIRLIQRTVAKVDIEIKQTNLINDINEQLPVYIGVVPARGGDFAPCLA
INFRLEINWIKTKDGVVPMVSPAFDLAYE"
complement(1334..1353)
/gene="SPAC343.01c"
/note="ctaactgagctcttgacag, splice branch and acceptor"

```

misc_feature      complement (1394..1399)
                  /gene="SPAC343.01c"
                  /note="gtaagt, splice donor sequence"
gene              2277..2825
                  /gene="SPAC343.02"
                  /note="SPAC343.02"
                  /note="SPAC343.02, len:183, SIMILARITY: Saccharomyces cerevisiae, YCT6 YEAST, ribosomal protein, mitochondrial, (169 aa), fasta scores: opt: 174, E():5.3e-05, (28.7% identity in 171 aa)"
                  /codon_start=1
                  /label=SPAC343.02
                  /product="mitochondrial ribosomal protein S19"
                  /protein_id="CAB52265.1"
                  /db_xref="GI:5706503"
                  /db_xref="SPRMBL:O9UT87"
                  /translation="MFNAKHFENLGGFQMLQKRGIGSLKRPYKPPVPGPKHNDVL
                  SLEKKCRSLDQSERPKMFRSQNRVAPVALVESKSPKSDVNRFGYLLR
                  IRRGPKSSILLNVMGVGEYLPIYSPOIKRIYVLKENGSLKRPRAVLSYLRQP
                  RFLPVSISLRKYIEQNHKP"
CDS              2484..2750
                  /gene="SPAC343.02"
                  /note="Match to P01245 Ribosomal_L19, Ribosomal protein L19 Score 24.79"
                  /join(3630..3648,3692..3814,4032..4174)
                  /gene="SPAC343.03"
                  /join(3630..3648,3692..3814,4032..4174)
                  /gene="SPAC343.03"
                  /note="SPAC343.03, len:95, SIMILARITY: Caenorhabditis elegans, Q20052, F35G12.9 PROTEIN, (135 aa), fasta scores, opt: 318, E():2e-14, (40% identity in 87 aa overlap)"
                  /codon_start=1
                  /label=SPAC343.03
                  /product="Highly conserved ring finger, similar to yeast apc11 SPC component"
                  /protein_id="CAB52266.1"
                  /db_xref="GI:5706504"
                  /db_xref="SPRMBL:O9UT86"
                  /translation="MKVKILRYHAIANMTMDPKDVCGICRVPDCCGQCTSPEDN
                  CPTVWCKKHIFIAHCTIQWLATSGSGCGCPMDRQTFVAVDSINKESTQ"
misc_feature      3649..3654
                  /gene="SPAC343.03"
                  /note="gtaagt, splice donor sequence"
misc_feature      3681..3691
                  /gene="SPAC343.03"
                  /note="caaccactag, splice branch and acceptor"
misc_feature      3815..3820
                  /gene="SPAC343.03"
                  /note="gtaaga, splice donor sequence"
misc_feature      4019..4031
                  /gene="SPAC343.03"
                  /note="caaccactag, splice branch and acceptor"
gene              complement (join(4455..5753,5878..6028,6083..6156))
                  /gene="SPAC343.04c"
                  /note="SPAC343.04c"
                  /note="SPAC343.04c, len:508, SIMILARITY: Caenorhabditis elegans, YZL1_CABE1, hypothetical 43.1 kd tnp-aap repeats containing protein K04g11.4 in chromosome x', (395 aa), fasta scores: opt: 424, E():1.1e-19, (28.6% identity in 301 aa)"
                  /codon_start=1
                  /label=SPAC343.04c
                  /product="WD repeat protein"
                  /protein_id="CAB52267.1"
                  /db_xref="GI:5706505"
                  /db_xref="SPRMBL:O9UT85"
                  /translation="MALDERPOLFVTHLLQRLNDGYDESLKALEKNGVSEFEDY
                  KRLQAVLOGDMITTAARFSLINQJDESKRKEQFLKQRCLELRSALICBAITVL
                  QNESTDFNKEKRLVSIILSNKSNNELITNGYGNTRLDLNLQSLAISYIEPLL
                  KRRLHLQADKYQVSQVYHNVLKNFSLDYKADPSELPTKEHYEHVHDSDEVQ
                  IYSHNGRYLASAKDTAIIIPVVALKEVRLIGHIDIVAVYRMSPPDRYILSCSD

```

```

misc_feature      complement (482..4601)
                  /gene="SPAC343.04c"
                  /note="Match to P000400 WD40, WD domain, G-beta repeat"
misc_feature      complement (5130..5228)
                  /gene="SPAC343.04c"
                  /note="Match to P000400 WD40, WD domain, G-beta repeat"
misc_feature      complement (5256..5360)
                  /gene="SPAC343.04c"
                  /note="Match to P000400 WD40, WD domain, G-beta repeat"
misc_feature      complement (5754..5769)
                  /gene="SPAC343.04c"
                  /note="ctaacatgctactctag, splice branch and acceptor"
misc_feature      complement (5872..5877)
                  /gene="SPAC343.04c"
                  /note="gtaaac, splice donor sequence"
misc_feature      complement (6029..6046)
                  /gene="SPAC343.04c"
                  /note="ctaacatgctactctag, splice branch and acceptor"
misc_feature      complement (6077..6082)
                  /gene="SPAC343.04c"
                  /note="gtaagt, splice donor sequence"
misc_feature      complement (6814..6884)
                  /product="tRNA-Gly"
                  /note="tRNA Gly anticodon GCC, Cove score 72.94"
                  /join(7781..7783,7839..7938,7981..9737)
                  /gene="SPAC343.05"
                  /note="synonym: ymal"
                  /join(7781..7783,7839..7938,7981..9737)
                  /gene="SPAC343.05"
                  /note="SPAC343.05, len:620"
                  /codon_start=1
                  /label=yml
                  /product="V-type ATPase; vacuolar ATP synthase catalytic subunit A"
                  /protein_id="CAB52268.1"

```

Alignment Scores:

Pred. No.:	Length:	42947
Score:	610.50	Matches: 152
Percent Similarity:	51.86%	Conservative: 71
Best Local Similarity:	35.35%	Mismatches: 140
Query Match:	27.38%	Indels: 67
DB:	8	Gaps: 13

US-10-069-062-7 (1-432) x SPAC343 (1-42947)

```

QY      6 SerAlaProGlyLysAlaPheLeuNlaGlyGlyLeuValLeuGluProIleTyrAsp 25
DB      1326 TCTGCTCGTGAAGTACTTATACCTGTGGATACATGTTTGATCTCAATACAC 1267
QY      26 AlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLysGlyThrSer 45
DB      1266 GGGCTTGTTATGTTTAACTGCAAGGATACGTTCCACACACACTTGGACAT--- 1210
QY      46 LeuYrGluSerArgIleLysIleSerSerProGlnPheAlaAsnGlyGluTrpGluTyr 65
DB      1209 ---AAATGGAACCGTTTCGTGTGAAGCCCTCAATTCATTAAACGTCGAATGCTTAT 1153
QY      66 HisIleSerSerAsnThrGluLysProArg-----GluVal 77
DB      1152 AAATATTGATTGACGTGATCTCTCAATTGAGTTCAAGATATATGAAAACGTAAATTG 1093
QY      78 GinSerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIle--- 96
DB      1092 GAAAAAATCCAAACCCCTGCTTCACTGGACACTTTTATGTAATTAATTAACCTTTT 1033
QY      97 -----GlnProThrGluAlaPheAspLeuGluIleIleIleTyrSerAspPro 112
DB      1032 TCACGGGGCGACAACTTATGCTGCAAGATCTTCAAGTTCAATTCAGTACGATATAT 973

```

QY 113 GlyTyrHisSerGlnGluAspThrGluThrLysSerSerAsnGlyGluLysThrPhe 132
 DB 972 GCGTATATACCAACACACATTAAGCTGATCAACAGACGATTCACAAATTCATTTTC 913
 QY 133 LeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAlaGly 152
 DB 912 TTA-----ACCTGACTTATAGGTCAAGTACAGATCGTTTGAAGCTCGGACGT 859
 QY 153 LeuValSerValAlaIleThrSerLeu-----SerHisPheIlePro----- 167
 DB 858 ATGATT-----ACAAAGCTTATAGGTTCTATTATTTTAAAGCTTGAAGCACTA 811
 QY 168 -----AsnValIleSerThrAsnLysAspIleLeuHis 178
 DB 810 ACAGATGACATGTCGACCAATCTCTCAAAATTTGATTCACACCAAGTCAATGATTC 751
 QY 179 AsnValAlaGlnIleAlaHisCysTyrAlaGlnLysLysIleGlySerGlyPheAspVal 198
 DB 750 ATCTGGGCACAAATTCGCCACTGTTCTGCACAAAGGAAAGTGAAGTGGTGTGATGTT 691
 QY 199 AlaThrAlaIleTyrGlyLeuIleValTyrArgArgPheGlnProAlaLeuIleAsnAsp 218
 DB 690 GCGCGTCCACTTGGGAAAGTTGATTATTAAGAGGTTGATCTTAATTAATTAACA 631
 QY 219 ValPheGlnValLeuGluSerAspProGlu-----LysPheProThrGluLeuLysLys 236
 DB 630 CTGCGTGTCCATACGACACAAATTAAGAAATTAATTTTCCACGCAATGCAAGAAA 571
 QY 237 LeuIleGluSerAsnTyrGluLysLysGluArgCysThrLeuProTyrGlyLysLys 256
 DB 570 ATAGTTTCTAAATAAGTGCAGAT-----GTTGTTCCATTTCAATTCATGCT 526
 QY 257 -----LeuIleMetGlyAspValLysGlyLysSerGluThrProLysLeuVal 272
 DB 555 GCTACTTATGTTCTTAATGGGAGCGTAGCTGGGACATGACACACCGGGATAGTA 466
 QY 273 SerArgValLeuGlnTyrLysLysGluLysProGluGluSerSerValValTyrAspGln 292
 DB 465 AAAAAAGTACAGCAATGCGAAGAAAGAAATCCAGAGAACTTAAGAATTCCTTATGAT 406
 QY 293 LeuAsnSerAlaAsnLeuGlnPheMetCysGluLeuArgGluMetArgGluLysTyrAsp 312
 DB 405 TTGATTTCAGAGCTCTTCC--ATCAAAACCTGTTCTTCTTCGAAAGCTTAGAC 349
 QY 313 SerAspProGluThrTyrIleLysGluLeuAspHisSerValGluProLeuThrValAla 332
 DB 348 TCTGAA-----CTTCAAAAGTCAA 331
 QY 333 IleLysAsnIleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGlu 352
 DB 330 TTTGCATCATTCGTCGATTTCTTCAAGAATTAATCCGTTGAAGCAAGGTTGACATGA 271
 QY 353 ProAspValGlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyGly 372
 DB 270 CCTTAAAGCAAAACAAATATTTTGAACAACATAGACAGTGGCTGCTGATGCTGTG 211
 QY 373 ValValProGlyAlaGlyGlyTyrAspAlaIleAlaValLeu----- 386
 DB 210 GGAGTTCTCGAGCGCGGATTTGATTCACAAATTTTGGCCATTAACCATACGAG 151
 QY 387 ValLeuGluAsnGlnValGlyAsnPhelys 396
 DB 150 ATAAATCGAAATGTTATCAAAACATGAAA 121
 RESULT 12
 AX441242 2396 bp DNA linear PAT 02-JUL-2002
 LOCUS AX441242
 DEFINITION Sequence 1 from Patent EP1209236.
 ACCESSION AX441242
 VERSION AX441242.1 GI:21690239
 KEYWORDS
 Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE
 1
 AUTHORS Meisner, R. and Lechelt-Kunze, C.
 TITLE Phosphomevalonate kinases from plants
 JOURNAL Patent: EP 1209236-A 1 29-MAY-2002;
 BAYER AG (DE)
 FEATURES
 source location/Qualifiers
 1..2396
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /db_xref="taxon:3702"
 BASE COUNT 669 a 492 c 558 g 677 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4,21e-31 Length: 2396
 Score: 475.00 Matches: 143
 Percent Similarity: 43.76% Conservative: 78
 Best Local Similarity: 28.32% Mismatches: 184
 Query Match: 21.30% Indels: 100
 DB: 6 Gaps: 15
 US-10-069-062-7 (1-432) x AX441242 (1-2396)
 QY 1 MetSerLysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyLysLeuValLeu 20
 DB 685 ATGGCTGTGTGCTCTTCCTGCTCCGGAAGTTTGGAGATGAGGCTACCTTGTACTC 744
 QY 21 GluProIleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThr 40
 DB 745 GAGAAAGCAATGCAAGGCTGTGTGATGACAAATGACACGTTTACGCAATTTGTAAG 804
 QY 41 ProLysGlyThrSerLeuLysGlu-----SerArgIleLysIle 53
 DB 805 CCAATCAAGCAAGAGTCAACCCGTAAGATTTGGGATGGAATGACAGATGCAATTA 864
 QY 54 SerSerProGlnPheAlaAsnGlyGluTyrGluTyrHisIleSerSerAsn-----Thr 71
 DB 865 ACATCAACACAGCTCTCGACA--GAAGCATGTATTAACCTGCACTGATCATTTGACT 921
 QY 72 GluLysProArgGluValGlnSerArgIleAsnProPheGlnAlaThrIlePheIle 91
 DB 922 CTTCAGTGTGTCTGCAAGTATTCAGAAACCCCTTGTGAGCATGCGATACAGTAT 981
 QY 92 ValLeuAlaTyrIleGlnProThrGluAlaPheAsp----- 103
 DB 982 GCTATAGCTGCTGCTCATTTGGCAACGAGAAAGCAAGAAATCATTCACAACTCTTA 1041
 QY 104 -----LeuGluIleIleIleTyrSerAspProGlyTyrHisSerGlnGluAsp--- 119
 DB 1042 TTGCAAGCTTGTGATTAACAATATTAGGCTCAATGACTTTTACTATATCGAACCAG 1101
 QY 119 ----- 119
 DB 1102 ATAAATCGCTGGGCTTCATTCACACCAAGATCCGCTGATCCCTTGCA 1161
 QY 120 -----ThrGluThrLysThrSerSerAsnGlyGluLysThrPheLeuTyrHisSerArg 137
 DB 1162 TCATTCACATTCANCGCGGAGTCAAAATGT-----GCTAAT 1200
 QY 138 AlaIleThrGluValGluLysThrGlyLeuGlySerSerAlaGlyLeuValSerValVal 157
 DB 1201 TCCAAGCTGAAGTGAAGAAACCTGGCTTATGATCTTCTGCAGCAATGACAAACAGCTGTG 1260
 QY 158 AlaThrSerLeuLeuSerHisPheIleProAsnValIleSer----- 171
 DB 1261 GTTCAGACTCTGTTA--CATTAATCT--GGAAGTGTGACCTATCTGATCCATGTAAA 1314
 QY 172 -----ThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHis 185


```

QY 137 ArgAlaIleThr-----GluValGluLysThrGlyLeuGlySer 149
DB 44253 CGGCTCATCCCGGTTATACGCCCTCTTGAGGAGGTTTCCAAAAGTGTCTAGGATCT 44312
QY 150 SerAlaGlyLeuValSerValAlaIleThrSerLeuSerHisPheIle----- 166
DB 44313 TCAGGGGCAATGACCACTAGTGTCTAGCGGTCTCCAACTTATCTTTTATTCG 44372
QY 167 -----ProteinValIleSer 171
DB 44373 AAGAAAAACATCCACATGATCCGTGAGAGGACCGGACCTACCTTACGTTTGAG 44432
QY 172 ThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysEtyrAlaGlnLysLys 191
DB 44433 GATGTAAACAGAGTTTATTCATGTATTAAGTCAAGTTGCCACCTGGGTACACAGGAAAA 44492
QY 192 IleGlySerGlyPheAspValAlaThrAlaIleTyrglyLeuIleValTyraArgPhe 211
DB 44493 GTGGGGAGCGGGTTGATGTCTACACGGCCACGTTCCGGACGTGTCTACCGACGGTTC 44552
QY 212 GlnProAlaLeuIleAsnAspValPheGlnValLeuGlnSerAspProGlnLysPhePro 231
DB 44553 TCCGACCGGCTCTGAAAAAGCTC-----GTAAAGGTAATGAACCTCCGAAGCGCGTC 44606
QY 232 Thr-----GluLeuLysLysLeuIleGlnSerAsn-----TrpGlnGluLysHisGln 247
DB 44607 ACTATTCCTCTTTACGTGAGTGTGTGCGAGACGATGAAAGTGGGTTCAAGAGATACC 44666
QY 248 ArgCysThrLeuProTyrglyIleLysLeuLeuMetGlyAspVal--LysGlyLysSer 266
DB 44667 ---TTCCGGCTTCCACCGGGTGTGACAGCTTCTTTTAAAGGAGTACACAAAGGAGAAACA 44723
QY 267 GlnThrProLysLeuValSerArgValLeuGlnIleProLysGluLysPheGlnLysSer 286
DB 44724 GAAACACCTGCATGCTGTGCAAGGTGATGATGAGGAGCGGTGCGTACACAGACCCC 44783
QY 287 SerValIleTyraAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGluLeuArgGlu 306
DB 44784 AACAGCTGTGGAGAGCACTTGCATGATGATAAATAATGTAAGGCGCTCCAGGGG 44843
QY 307 MetAlaGluLysTyraSerSerAspProGlnLysTyra-----IleLysGluLeu-- 322
DB 44844 CTAAATCAAAAGCTCTCAGGAGGCGCGGTGCTATCACCGACGAGCTGAAAAATTTGAG 44903
QY 323 -----AspHisSer-----ValGlnPro 328
DB 44904 AGCGTGTATTTGGCTAAACACACACCCCTCCACGAGGCTGAGAGGCTATGGGTGAGGCG 44963
QY 329 LeuThrValAlaIleLysAsnIleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGlu 348
DB 44964 GCGTCAAGTTGCC---TCAACAGAGTGGCGCTTACCTTAAGGAGATGGAGAGCGGCTCAG 45020
QY 349 ValProIleGlnProAspValAlaGlnThrLeuLeuAspArgCysGlnGlnLysLeuProGly 368
DB 45021 GTGCAAAATTAAGACCTCCGGAACCTCACTCTTGAATGACCAAGTGTCTATTCGCCGT 45080
QY 369 CysValIleGlyValValAlaProGlyAlaGlyLysTyraAspAlaIleAlaValLeuValLeu 388
DB 45081 GTGTTTGGCGGTGGTGGCTGGCGCTGGCGGATGACGCGGTGTTGGCTTATGTTCTC 45140
QY 389 GluAsnGlnVal 392
DB 45141 GCGGAGAGAGTT 45152

```

RESULT 14
 LOCUS AF429385 1785 bp mRNA linear PLN 25-OCT-2001
 DEFINITION Hevea brasiliensis phosphomevalonate kinase mRNA, complete cds.
 ACCESSION AF429385
 VERSION AF429385.1 GI:16417947
 KEYWORDS
 SOURCE Hevea brasiliensis (Para rubber tree)

```

ORGANISM Hevea brasiliensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Malpighiales; Euphorbiaceae; Hevea.
REFERENCE 1 (bases 1 to 1785)
AUTHORS Hallahan,D.L. and Keiper-Hrynko,N.M.
TITLE Genes involved in the biosynthesis of isopentenyl diphosphate in
the rubber tree Hevea brasiliensis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1785)
AUTHORS Keiper-Hrynko,N.M. and Hallahan,D.L.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2001) Central Research and Development, E. I.
Dumont de Nemours Co., Wilmington, DE 19800, USA
FEATURES
source location/Qualifiers
1..1785
/organism="Hevea brasiliensis"
/mol_type="mRNA"
/cultivar="RRIM600"
/db_xref="taxon:3981"
/cell_type="latex"
121..1632
/codon_start=1
/product="phosphomevalonate kinase"
/protein_id="AA18926.1"
/db_xref="GI:16417948"
/translation="MAVVASAPGVLTMTGVTLLERPNAGVILSTNAPFAIVPRIVD
EIKRDSNANMTDVKLTSPOLARSILYKLSKINLALQCVSSASARNPVEAVOPAVA
AAHATLDDKKNVINKLQGLDITITGNDPISYRNEICGAPLTPESIALAPFS
SITNPEKRNQONCKPEVAKTGLSSSAMTAFAVVALHLHGLVDLSSCKEKESDL
DLVHIQTAHICIAOGKVGSGFDVSAVYGSRRYRSPVLSAODAGKGIPLQEV
SNLKGKWDHRTNFSLEPLMSLLGEGTGSSTPSVAGALKKMKODTOKSGDTR
KLSEANSALTEOFNLSILAEHMDAYKCVISDSTSKSEKVIQDAPRBAVAVKAL
LGSNANMCIPTNWRQMGAEAGVLEPESQRIPLDTTMMGCVILAGPAGPAPAV
AVTIGDGSNTNAAKWSLNTVALIYREDPNVLLSEGDPRTEITTAFAVAVHT"
BASE COUNT 554 a 351 c 406 g 474 t
ORIGIN
Alignment Scores:
Pred. No.: 9.74e-29 Length: 1785
Score: 445.50 Matches: 138
Percent Similarity: 44.84% Conservative: 88
Best Local Similarity: 27.38% Mismatches: 179
Query Match: 19.98% Indels: 99
DB: 8 Gaps: 16
US-10-069-062-7 (1-432) x AF429385 (1-1785)
QY 1 MetSerLysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyLysTyraLeuValLeu 20
DB 121 ATGGCTGTAGTGGCTCTGCTCCGGGTAGGCTGTGATGATGAGTGGGCTTACTCATATTG 180
QY 21 GlnProIleTyraAspAlaTyraValThrAlaLeuSerSerArgMetHisAlaValIleThr 40
DB 181 GAAAGACCAATGACGAGGATGTGACTGACACAATGTGATCTATGCACTATGTGAAG 240
QY 41 ProLysGlyThrSerLeuLysGlu-----SerArgIleLysIle 53
DB 241 CCTATTTAAGATGAATTAACCTGATGATGGGACATGGGACATGACATGATGATAATTA 300
QY 54 SerSerProGlnPheAlaAsnGlyLysLysTyraGluLysLysSerSerAsnThrGluLys 73
DB 301 ACATCTCCCAACATGACAGG---GAAAGCTTGTACAAATGTGTCACTGAAAAATTGACT 357
QY 74 ProArgGluValGlnSerArgIle-----AsnProPheLeuGlnAlaThrIlePheIle 91
DB 358 CTTCAGTGTGTCTCTTCAAGTGCATCAAGGAACCATTTGTGGAACAAGCAGTGCATTTT 417
QY 92 ValLeuAlaTyraIleGlnProThrGlnAlaPheSer----- 103
DB 418 GCTGTAGACGTGCACATGCAACCTTGACAAATATTAAGAAATGCTTTAAACAGCTA 477
QY 104 -----LeuGlnIleIleIleTyraSerAspProGlyLysLysSerGlnLysPhe 119

```

```

Db      478 CTCTGCAAGCTCTGATATTAACAATATAGTACCAATGACTTATTCATACCGAAT 537
QY      119 -----
Db      538 GAGATTGAAGCATGAGACTCCCTTGACACGAAATCATGTGCTGACTCTCTCTTT 597
QY      120 -----ThrgluthrlyserSerSenglyglubserPheleuThryHisser 136
Db      598 TCTCATATACCTTCATATGATAGAGAGCAATATGACAA----- 636
QY      137 ArgAlaIlethrGluValGluYstrnglyleuGlySerSerAlaGlyLeuValSerVal 156
Db      637 AACTGCAAGCTGAGTAGTAAACTGATGTTGATTCAGACAGCAATGACCACTGCT 696
QY      157 ValAlaIlethrSerleuLeuSerHisPhe---IleProAsnValIleSerThrAsnlys--- 174
Db      697 GAGTTGCTGCTTACTCATCATCTGATGATGGTGGATCTTTCATCTCTGTAAAGAG 756
QY      175 -----AspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
Db      757 AAGAAATTTTCGATCTGATTTGTTGATCATATATAGCCCAACTGCTCCATTTGATGCA 816
QY      189 GlnlyValSerIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
Db      817 CAAGGAAAGTCGCGAGTGAATTTGATGTTGATTCGACAGTTATGCGAGTCATCATGATAC 876
QY      209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGlnLeuSerAspProGlu 228
Db      877 GTGCGCTTCTCCACAAAGTCTTCTCTGCTGACAGATGCTGGAGAAAGAAATTCAC--- 933
QY      229 LysPheProThrGlnLeuLysLysLeuIleGlySerAsnTrpGlnGlnLysHisGluArg 248
Db      934 ----TTACAGAGAGTCAATTTCTAATCCCTAAAGAAATGGACCATGAGAGACTAG 590
QY      249 CysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspVal---LysGlyGlySerGlu 267
Db      991 TTTTCTCTCCACCATGATGAGCCCTGACTAGAGACGACAGCAAGTGAAGATCTTCC 1050
QY      268 ThrProLysLeuValSerArgValLeuGlnTrpLysGlyGlnLysProGlnGlnLysSer 287
Db      1051 AGCCCACTCAATGCTGCTCTAAAGAAATGCGAAGCTGATACCTGAATATCCCA 1110
QY      288 ValValTyrAspGlnLeuLeuSerAlaAsn-----LeuGlnPheMet 301
Db      1111 GAAACATGAGAGAAAGTTGTCAGAGGCAATTCAGCACTTGAACGCAATTCATATTTTA 1170
QY      302 LysGlnLeuArgGluMetArgGluLysTyrAspSer----- 313
Db      1171 AGCAAGCTCGCAGAA-----GAAACATGAGAGCGCTATAAATGTGTATAGACAGTTGC 1224
QY      314 AspProGlnThrTyrIleLysGlnLeuAspHisSerValGlnProLeuThrValAlaIle 333
Db      1225 AGCACAAAAACTCAGAGAAAGTGAATGAGCAGGCACTGAAACCCGAGAGAAAGCAAGTT 1284
QY      334 -----LysAsnIleArgPheGlyLeuGlnAla 342
Db      1285 GTTAAAGCATATATAGATCAAGAAATGCCATCTCAGATCAAGAAATTCATATCCCGCAG 1344
QY      343 LeuThrGlnLysSerGlnValProIleGlnProAspValGlnThrGlnLeuLeuAspArg 362
Db      1345 ATGGGAGAGGCTGACAGGTGTTCCGATAGAGCTGATTCACAGACTGATTTGGATACT 1404
QY      363 CysGlnGlnIleProGlyCysValGlyGlyValValProGlyValAlaGlyIleTyrAspAla 382
Db      1405 ACTATGAATATATGAGTGAAGCTTGTGGCTGAGATTCGAGAGCTGGTGGTTATGAGA 1464
QY      383 IleAlaValIleValLeuGlnLeuGlnValGlyAsnPheLysGln-----LysThrLeu 400
Db      1465 GTCTTCGCTTACCTTATAGGAGCTGTGATACCAATGTGGCAAAAGCTTGGAGTTACTC 1524
QY      401 GluAsnProAspTyrPheHisAsnValTyrTrpValAspLeuGlnGlnThrGlnGly 420

```

```

Db      1525 -----AATGTTCTGCGCTGTGTGTTAGAGAAAGACCTTAATGCT 1563
QY      421 ValLeuGlnGlu 424
Db      1564 GTTTGTTAGAA 1575

RESULT 15
LMFLCHR15_1/c
WPCOMMENT
Sequence split into 8 fragments
Fragment Name      Begin      End
LMFLCHR15_0      1          110000
LMFLCHR15_1      100001     210000
LMFLCHR15_2      200001     310000
LMFLCHR15_3      300001     410000
LMFLCHR15_4      400001     510000
LMFLCHR15_5      500001     610000
LMFLCHR15_6      600001     710000
LMFLCHR15_7      700001     720704
Continuation (2 of 8) of LMFLCHR15 from base 100001 (AL160371.1elehmania major chromosome

Alignment Scores:
Pred. No.:      9,88e-21      Length:      110000
Score:          379.50      Matches:      118
Percent Similarity: 43.56%      Conservative: 68
Best Local Similarity: 27.63%      Mismatches: 188
Query Match:    17.02%      Indels:      53
DB:              2          Gaps:      15

US-10-069-062-7 (1-432) x LMFLCHR15_1 (1-110000)
QY      6 SerAlaProGlyLysAlaPheLeuAlaGlyIleTyrLeuValLeuGlnProIleTyrAsp 25
Db      82535 TCAGCACCCGCGAAGTTCTATCTTGAGTACCTTATGTGAGTCGACACGCG 82476
QY      26 AlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys----- 42
Db      82475 GCCAATGTC---GGTATCTCGATCCGCGCTGAACGCGCCCTTCACACGCGCATCTCAAG 82419
QY      43 -----GlyThrSerLeuLysGluSerArgIleLysIleSerPro 56
Db      82418 GCGGAGCCCAACAGACCGCGCCCTTCG---GGANAAAGCATATTCACATGAGTGGCCG 82362
QY      57 GlnPheAlaAsnGlyIleTrpGlnLysIleSerSerAsnThrGlnLysAspProArgGlu 76
Db      82361 CAGTTCCGTCAG-----TCATTCTGCTTCCTGCCAATACCTCAGCGCGCCGACCC 82311
QY      77 Val-----GlnSerArgIleAsnProPheLeuGlnAlaThrIlePheIle 91
Db      82310 GTCCGCGTGAANCANACNAGAGGAGGAGGANNCCGTATATCTTCACGNTGCTCTGTAN 82251
QY      92 ValLeuAlaTyrIleGln-----ProThrGlnAlaPheAspLeuGlnIleIle 107
Db      82250 TCTGTGGGCGCGCCGACCTGCTGCGGACAGCAGCACCGATGGC---GAATTTGGGTAGAG 82194
QY      108 IleTyrSerAspProGlyTyrHisSerGlnGlnIleAspThrGlnLysThrLysSerSerAsn 127
Db      82193 CTTCTGCGTCGAACAAGATCTTATACGTCAACCCANNTATNTANAGAGTCANANAGAGAC 82134
QY      128 -----GlyGlnLysThrPheLeuTyrHisSerArgAlaIleThrGlnValGln 143
Db      82133 GTGNAAGTGGCCAAACNTGCGACGACCCGACCCGACGACGTCGTCGTCGTCATTC 82074
QY      144 LysThrGlnLysGlySerSerAlaGlyLeuValSerValAlaIleThrSerLeuLeuSer 163
Db      82073 AANACCGCTGGGATCTTCTGCGCGCATGACNACNACATCGATGGCGCTGCTNTGCCAC 82014
QY      164 HisPheIleProAsnValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIle 183
Db      82013 CATTTN-----AAGCGGAGTGGGTGCTCACACAGTACGTCATTCGATCCGACAGATC 81960
QY      184 AlaHisCysTyrAlaGlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyr 203

```

[illegible]

FEATURES	Source
misc_feature	1..1037 /organism="Saccharomyces bayanus" /mol_type="genomic DNA" /strain="ClIB 533" /variety="uvatum" /db_xref="taxon:4931" /clone="AS0A025E12" /clone_11b="AS0AA" /note="end : 17" <3..>560 /note="similar to Saccharomyces cerevisiae ORF YMR220w [EGG8 ; phosphomevalonate kinase]" /evidence=not experimental complement(<788..>1005) /note="similar to Saccharomyces cerevisiae ORF YMR221c [weak similarity to photosystem II protein D2] 1 putative frameshift (s)" /evidence=not experimental
misc_feature	161 c 214 g 298 t 1 others
BASE COUNT	363 a 161 c 214 g 298 t 1 others
ORIGIN	
Alignment Scores:	
Pred. No.:	1, 91e-15 Length: 1037
Score:	287.00 Matches: 71
Percent Similarity:	55.85% Conservative: 34
Best Local Similarity:	37.77% Mismatches: 65
Query Match:	12.87% Indels: 18
DB:	11 Gaps: 5
US-10-069-062-7 (1-432) x CNS06037 (1-1037)	
Qy	257 LeuLeuMeGlyAspValIyGlyGlySerGluThrProIyLeuValSerArgValIleu 276
Db	3 TTATGATGGGCGATATTAAAGTGTTCAAGAAACGGTTAAACTGTTCAAAAGGTGAAA 62
Qy	277 GlnThrIyLeuGlyIyPProGlnIySerSerValValIyTryAspGlnLeuAsnSerAla 296
Db	63 AAATGGTACGATTCACACTTACACAAAGAGCTTGAGAGATATACCGAGCTGGATCAGCCT 122
Qy	297 AsnLeuGlnPheMetIyGlyLeuIyArgGluIyMetArgGluIyTryAspSerAspProGlu 316
Db	123 AATTCAAGGTTATATGAGATGAACTATCTAACTAAACGGCTTATATTACTCTCAT---GAT 179
Qy	317 ThrTyrlIyLeuGlyIyLeuAspHisSerValGlu----- 327
Db	180 TATTATAGTGGCAATATTGAGGCTATAGAAAAAGATGACTTACTTGCCAAAGTAT 239
Qy	328 -----ProLeuThrValAlaIleIyAsnIleArgIyGlyLeuGlnAlaIleu 343
Db	240 TCAGAAATCACAGACGTGAGAGATTCTCTGTACACGATTAGACGTTGTTCAAGAAAAAT 299

OY		344	ThglnlyseScGlvalProlleGluProAspValGlnThnGlnLeuLysAspArgCys	363
Dd		300	ACTATAGATCTGGAGCCGATATCAACACCCTGTACAGACCACAAGTATTGGACGATTGT	359
OY		364	GInguilleProcglCysValGlyGlyAlaValValProGlyAlaGlyGlyTYrAspAlalle	383
Dd		360	CAGACTTTAGAAGGTGTTCTCACATCGCTTGAAACTGGTGCTGGTATTGATGCCATT	419
OY		384	AlaValLeuValLeuGluLysGlnValGlyIleAsnPhelyrgGlnYrThrLeuGluAsnPro	403
Dd		420	GCAGTATTATGCC-----ACACAGGATGTCGACCTCAAAACCAAACACT---GCCGGTAT	470
OY		404	AspyrPrphHisAsnValTYrTrpValAspLeuGluGlnInThnClnglVyaLeuGlu	423
Dd		471	AAAAGATTTCTTAATGTTCAATGGTTTAGACGCTCTGTCAAGCTGATTGGGGTATTGAATAA	530
OY		424	Gltys---ProGluAspTYrile	430
Dd		531	GAAGAAAGAACCGAGACATATCTTT	554
RESULT 17				
LOCUS	AC079041/C	119420 bp	DNA	linear PLN 19-JAN-2001
DEFINITION	Arabidopsis thaliana chromosome 1 BAC F5M6 genomic sequence,			
ACCESSION	AC079041			
VERSION	AC079041.4	GI:10086525		
KEYWORDS	HTG.			
SOURCE	Arabidopsis thaliana (thale cress)			
ORGANISM	Arabidopsis thaliana			
REFERENCE				
AUTHORS	Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J., Wu,D., Maiti,R., Roming,C.M., Koo,H., Fujii,C.Y., Uteback,T.R., Barnstead,M.E., Bowman,C.L., White,O., Niernan,W.C., and Fraser,C.M.			
JOURNAL	Unpublished			
TITLE	2 (bases 1 to 119420)			
REFERENCE	Town,C.D. and Kaul,S.			
AUTHORS	Submitted (17-AUG-2000) The Institute for Genomic Research, 9712			
JOURNAL	Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org			
TITLE	3 (bases 1 to 119420)			
AUTHORS	Town,C.D. and Kaul,S.			
JOURNAL	Direct Submission			
TITLE	Submitted (19-JAN-2001) The Institute for Genomic Research, 9712			
REFERENCE	Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org			
AUTHORS	On Sep 12, 2000 this sequence version replaced gi:19451516.			
JOURNAL	Address all correspondence to:atet@tigr.org			
COMMENT				

BAC clone F5M6 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from Sp6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
<http://CCB-081.mit.edu/GENSCAN.html>), GenemarkMM (Mark Borodovsky,
<http://genemark.biology.gatech.edu/genemark/>), GlimmerA (a variant
of GlimmerW, see Mihnea Perlea,
<http://www.tigr.org/soflab/glimmer.htm#glimmer.html>, and
GeneSplicer (Mihnea Perlea and Steven Salzberg, contact
mperlea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(<http://www.tigr.org/tdb/cgi.shtml>). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).

Simple repeats are identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).

Location/Qualifiers

1. .119420

/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/culivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="FSM6"

misc_feature

1

/note="23582 nt beyond this point were not included in the submitted sequence due to an overlap with another BAC (T12021) - the overlap is larger than 23582, but due to lack of bac end sequences, we cannot be sure where the overlap ends."
complement(2532..2631)
/rpt_family="AT_rich"
complement(2799..2818)
/rpt_family="(TA)n"
complement(3149..3184)
/rpt_family="AT_rich"
complement(3201..3256)
/rpt_family="AT_rich"
complement(3470..3506)
/rpt_family="AT_rich"
complement(3585..3630)
/rpt_family="AT_rich"
complement(4364..4393)
/rpt_family="AT_rich"
complement(6251..9220)
/gene="FSM6.3"
/note="similar to p68 RNA helicase (Schizosaccharomyces pombe) GI:173419"
complement(join(6251..6529,6641..6733,6838..7002,7098..7191,7285..7353,7437..7504,7587..7643,7723..7913..7991,8146..8226,8316..8423,8522..8652,8754..>9220)
/gene="FSM6.3"
complement(join(6434..6529,6641..6733,6838..7002,7098..7191,7285..7353,7437..7504,7587..7643,7723..7913..7991,8146..8226,8316..8423,8522..8652,8754..9220)
/gene="FSM6.3"
/codon_start=1
/product="p68 RNA helicase, putative"
/protein_id="AAG50723.1"
/db_xref="GI:12311302"
/translation="MAGOKOELPVSGEPLAVESPMTNKKKKSKNKHTEENHEEYPOETVTEBELSNKEKKKKRKEKSEKKNKQDVBEKLEADLDEGSEEOKKVYTKGVSEAKYALATTPRBSNLPENYVDCCTFKPEPSYIQSTWTFPLDGRLLIATKSGKTLTAFGIPALHVLNKKKIKGGSKKNVPLTVLSPTRELAQVLSDVAREGEPCGLSKICVYGSGKGPQISARSGVDIIGTPGRRLDIESTVELASVSPFVLADEBRLMDJGPEEPRFRLISNTNKYRMWFSATWPLDVHKLAEQMDENPIKIIIGSDVLAANHDWQIEVLDERADORLIALLEKHKSKQKRVLVALYKVAERLERLQORGMKAVSHGKKAQSERRSLSLFKEGSCPLVATDVAAKGLDIPDVEVINYNTPLTDEYVHRIGTARGAKGVAAHTFTPLNKSLAGELVNLRAAGVVPADLLFGLTHKKESKLYAHFKETIAAAPKATKTFPNDSDDE"
complement(8985..9154)
/rpt_family="(GAA)n"
9487..10008
/gene="FSM6.4"
/note="predicted by genemark hmm"
<9487..>10008
/gene="FSM6.4"
9487..10008
/gene="FSM6.4"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAG50726.1"
/db_xref="GI:12311305"
/translation="MDAKEDVPRTPGARGGVATKASHVITVSNMSVROPLDYKTTMNLNPLSLDPIYDNDONTLIKDDCGRVKDKGIYDAVFTNGRKMAMLYKHNRASYLNNTFSIRKLPDAGVKGSGNVGRFHSVETSNVTTISVANDFTNNKGGDDGVNTVGTIIRARSTNDSDN"

gene complement(10947. .13334)
/gene="FSM6.5"
/note="contains Pfam profile: PF01397: Terpene synthase family"

mRNA complement(join(<10947. .11246,11336. .11584,11696. .11828, 11920. .12138,12229. .12613,12713. .12983,13071. .>13334))
/gene="FSM6.5"
complement(join(10947. .11246,11336. .11584,11696. .11828, 11920. .12138,12229. .12613,12713. .12983,13071. .13334))
/gene="FSM6.5"
/codon_start=1
/product="terpene synthase family protein, putative"
/protein_id="AAG50729.1"
/db_xref="GI:12321308"
/translation="MEAIRIGFGLPMVHVSVPICLTTRTCRPPRORLLSHSTPSWPKAK ODDPFASSSSKSSDDKEKIRLIHLVSMGISTYHPDKELQDLKISFTTDLDTITGEDDLLE TISIMEVFKLIGHKMSCDAPFRGNDREFESLVDRFMQLQLEVAHLTPCEVI MDHALSTRNHLSTSGNASTASPHLLHIGNSLYIPRCNIEVLAAREVLSYEOB EKHDELILKPAKLNFQCFHYOELKTLTKMWRDIDLSKLPYIRDLVESHVLAAG PYEPHYSLGRILIVAKIMIMVVDVDTYAVATLPQVKALTECLQSIKESDLPLYLR IYGLSLFDVWGIIEBREMPLGRLYRKVOVEKIKITTKAYOSIARKARGHSTDEY MKGVLTAGMADYAAACPTIGMEDINEKAFEWLNSPLIILKLTAMFRANDVGTET EINRGSEVANGLNCYKQYGVTEKASRELKMTVYRKVVVEEFMNSHDHVRQVLLR CLNIRAFIDVFYTEGDGSESPKGI EHFMTSLYLHPILPS"

repeat_region complement(14658. .14692)
/rpt_family="AT_rich"

gene /gene="FSM6.6"
join(<18323. .18520,18618. .19016)
join(<18323. .18520,18618. .18848)
/gene="FSM6.6"
/codon_start=1
/product="unknown protein"
/protein_id="AAG50708.1"
/db_xref="GI:12321287"
/translation="MSYFPQDDKEVLAAPNSDIVAATTTATATGAGIEVATFEKVEH PVEPLNDLPICPLPEPSILNDGRVIAKMRVSGMRKGDQIVDDEAASEDSGAPK PRPPQPNRSILPSLSAPBNHLINLLEECNAMPVTSKNG"

CDS complement(20180. .24092)
/gene="FSM6.7"
/note="contains Pfam profile: PF00503: G protein alpha subunit"

mRNA complement(join(<20180. .20774,20852. .21024,21105. .21361, 21443. .21617,21696. .21875,21963. .22067,22155. .22292, 22386. .23390,23481. .23566,23944. .24092))
/gene="FSM6.7"
complement(join(20180. .20774,20852. .21024,21105. .21361, 21443. .21617,21696. .21875,21963. .22067,22155. .22292, 22386. .23309))
/gene="FSM6.7"
/codon_start=1
/product="G-protein, putative"
/protein_id="AAG50710.1"
/db_xref="GI:12321289"
/translation="MEKKDSEBSKEMVRKMLPGAPLPEDPSEFVSIALEYTGGP VHDIPRSPVDVNPVNNPIPLPVSRIGRSGSSGSPASESVSYLAHNPPSSSG SASVSPVSGHRONGNQRPPVYKRPVDHRIIGREAEEDNNVEAREKRVHE CTASTRKRRKKKSSCYRCGAKMENKEKTCVDEKYGCNCLVLRAMGMPGRKVCV IGQALDESKRSKLGHSYRLSLSPLEVKQIMAKKECTANQLPEBQILSPVGLPK BEMADLNLCLLPQKLKPGRYWYDKSGIWMGEGEKPRVLSNINFGKSLKNGVLSNG NTEVINGREITKLELILKLANVOCPPDTHWVYDDGRVEEBOGNNIRGNIMEAST RPMCALPSLPVPOGCGTVOSSSNVAPVNYIEHKOKILLIGLIGSSGSTITKOA KFLVKNKSVLELDIKLAVOSNTRYISLIDGEREERBEELSTRGLNVEGSGG BEANDEGVITTPQSVYTLNPLRKHPSDWLIDLIATGDIDAPPAATRYAPLVEEVMK DPALQATYRRKDELHFLPDVAEYFLSRAMEVSNSEYSESRIVAAEVTGNGALFM EBSLSDHSPMSSEYBENDALSSPOKQLIRVNAKGKSNKVMEMFEDRAVIFCI LSADYDQINTEPSSGTVQYOKIMLOSKELEFSMWKHCFCQDTPFILLINLYDOEEK LNRAPLSCDMSWSDRCVPRTNNNOSIAVOAYFYAMAKFLPSTITQCKLFWOAKA RDRANVDEGFKYVREVLKDEKESYIANGGGEFSYSTDMSSSFYREE"

repeat_region complement(24371. .24395)
/rpt_family="AT_rich"

gene 25457. .27277
/gene="FSM6.8"
/note="contains multiple FPR domains: PF01535: PPR repeat"

Alignment Scores:
Prid. No.: 1,526-11 Length: 119420
Score: 273.00 Matches: 144
Percent Similarity: 28.11% Conservative: 89
Best Local Similarity: 17.37% Mismatches: 166
Query Match: 12.24% Indels: 434
DB: 8 Gaps: 18

US-10-069-062-7 (1-432) x AC079041 (1-119420)

QY 6 SerAlaProGluValAlaPheLeuLagIyGlyYrLeuValLeuGluProIleYrAsp 25
|||||
Db 29974 TCTGCTCCTGGAAAGCTTTGTGACCTGAGGCTACCTTGATCTGAGAAAGCCAAATGCA 29915
|||||
QY 26 AlaYrValThrAlaLeuSerSerArgMetHisAlaValIleThrProIleGlyThrSer 45
|||||
Db 29914 GGGCTTGCTGAGTACAAATGCAACGTTTACGCGATTTGCAACCAATCAAGAAAGA 29855
|||||
QY 46 LeuIyGlu----- 48
:::
Db 29854 GTCAAGCCTGAAGCTTGCGACATGGGTATACCTCTTCATTTATGTATGGGA 29795
48 ----- 48
QY 48 ----- 48
Db 29794 GTAATATGTTGACAGCATTTAGAGGGTTTCATGTTGTGTGAGACAATTTGTTGTTT 29735
49 ----- 60
QY 49 ----- 60
Db 29734 GTTTCCTATTGTCGAGAAATGACAGATGCAATTAATCAATCCACAGCTTCGAGA 29675
61 GlyIuTrpGluYrHisIleSerSerArgThrGluYrProArgGluValGlnSerArg 80
|||||
Db 29674 ---GAAGCAGTATAAAGTCACTGATCAATGATTTGACTCTTCACTGCTGTGTCGAAAG 29618
|||||
QY 81 IleAsnPro-----PheLeuGluLalThrIlePheIleValLeuAla----- 94
:::
Db 29617 CAAGATCGAGAGACTTCTTCAGATGCGCTGC-ATATTTTTTTTTTATCTCGATATAT 29559
94 ----- 94
QY 94 ----- 94
Db 29558 AAGATTTTTTTTCATATCAAGACGTCTGTGAGATTTGATTTACAATATTAATGCA 29499
94 ----- 94
QY 94 ----- 94
Db 29498 TGAAGCACTTACACTATTTCTTTTCTTAATAAATGAAGTATTCAGAAACCCCT 29439
94 ----- 94
QY 94 ----- 94
Db 29438 TTGAGACATGCAATACAGATGATGCTATAGCTGCTCATTTGGCAACGAGAGACA 29379
94 ----- 94
QY 94 ----- 94
Db 29378 AAGAATCATTCACAACCTTTATTCAGAGCAATCTTCAAGTCTCTCTCTTTTATA 29319
94 ----- 94
QY 94 ----- 94
Db 29318 ATCATTTTGTTCATGAAGAAGATGTATGATGAACATTTCTTAAATGTGTGTTG 29259
94 ----- 94
QY 94 ----- 94
Db 29258 CAGGCTTGATATAACAATATATAGGCTTCAATGACTTTACTCATATGGAACAGAGTTT 29199
94 ----- 94
QY 94 ----- 94
Db 29198 TCTATCTTTTCTTGTGGAACCTGTCAAGTGCATTTGATTTGTAATTTCTTT 29139
95 ---TyrIleGlnProThrGluAlaPhe-----AspLeuGluIleIleIleYrSerAsp 111
|||||
Db 29138 GATATTCAAATCCTTTTGAGATGTTGTACAAATATGTGATGTTCTTGTATATTTCAAT 29079

gene /organism="Oceanobacillus iheyensis HTE831"
/mol_type="genomic DNA"
/strain="HTE831"
/db_xref="taxon:221109"
/note="halotolerant and alkaliphilic"
300..1643
/gene="dnaa"
/gene="dnaa"
/note="CDS ID OB0001"
/codon_start=1
/transl_table=11
/product="chromosomal replication initiation protein"
/protein_id="BAC11957.1"
/db_xref="GI:22775679"
/translation="MENIELMSATLTKIEELSKSPDPYMLKNTKAELEKDTLLIS
ANEFAPDLEENQYNTLISOMLEVTGSELTNKFPIPSLEIEBQKPKRQSTDT
GSPKMSKSTPTPTFYGAGNRPAHMSLVAENPAKATNPLTYGGVGLKTHM
HAIGHVRDHPNNAKVYLTSEKFTNFMAMDNKSNFRNRYNIDVLLIDDIQFI
AKESIOEFTFTFALHGESKQIISDRPKKPTLEDRLSRPEWGLIDITPDP
LETRIALINKKAKAEGLDIPNEMLIYINQINTNIEGALIRVAASLYNODIP
SLAADLKDIIPSSKPKETIPALOEIVAEYHRLDEPAKRTKSAFFROIAMYL
SEELDASPCKIGERPGRDHTVYHAHEKISKMLEQDTELDRLDIELKELKLSI"
1830..2966
/gene="dnaN"
/gene="dnaN"
1830..2966
/note="CDS ID OB0002"
/BC_number="2,7,7,7"
/codon_start=1
/transl_table=11
/product="DNA-directed DNA polymerase III beta chain"
/protein_id="BAC11958.1"
/db_xref="GI:22775680"
/translation="WRFTIQDKLINGSNWKAKISARVPIPLTGMKLEVNAGVTL
TSDSISIEYIPIBEDGIVVENIEBGTIILQAKYPPDIVRKLPESTVDIVDDQL
NRTITSGAEFNLNGQSAEYVQLPVPVQTEENFELPIDKMKIQVFAVSTMETPR
ILTVNLKLVNDSLFTATDSHRLAREIPVSNAPILSIVAPKSLNKLKIGDS
BETVLSVTNNQIILFRTKLANFLSLDLNGYETSLRIPESKTKIOLKTKELLTID
BSLAKERRNNVKNAPNGSMITSISSNPEVGNVEITADOMEGBEVKISPSKY
MIDALKAIEHYDVQIEFTGAMPFTIIRPVGDSITIIQLIPVTTY"
3219..3440
/gene="OB0003"
3219..3440
/gene="OB0003"
/note="CDS ID OB0003"
/codon_start=1
/transl_table=11
/product="hypothetical conserved protein"
/protein_id="BAC11959.1"
/db_xref="GI:22775681"
/translation="MHEQIQIDTEYITLGGQILKLNFLSESGMVKTFLOEBGALVNGH
LEQRGRKLYPDDVVEIQGISYIVKED"
3443..4552
/gene="recF"
/gene="recF"
3443..4552
/gene="recF"
/note="CDS ID OB0004"
/codon_start=1
/transl_table=11
/product="DNA repair and genetic recombination"
/protein_id="BAC11960.1"
/db_xref="GI:22775682"
/translation="WHIKELTLNRYNDQLEIAPDDQINVTIGENAGKTNMAYI
VLSRSHRTPREKELIQNDKYAKIBGTRKNSIPQISITSKGKAKVNHLEOR
RLSDYISVNVVFAPEDDITVKGAPQIRRRPMDELQIQETLYIYHAAQYQVLROR
NHLIKQLRKNSDITMLLEVLTDQLEHASILLEBRFTYLELRMAQPIRHGISREL
EQLLEIYSPSIESEDANKKIGNIYQKFAVKQETLERGTTLAGPRDDILFPVNG
KDVQYSGOQORTVALSIKLAIEILIOEYVEYPIILLDDVLSLDDYRQSHLNTI
QKQVQTFSTTSVEGIIHETLQOALELFRVTDGVNA"
4588..4869
/gene="OB0005"
4588..4869

gene /gene="OB0005"
/note="CDS ID OB0005"
/codon_start=1
/transl_table=11
/product="hypothetical conserved protein"
/protein_id="BAC11961.1"
/db_xref="GI:22775683"
/translation="MRIHNGNVIRTEKIVAILIDCNLSSSIYDENMBAWKNKV
SGPRHAKSIMLTDHYVFSSLVATLKRSMSMSTINKLDDYSBELSI"
5109..7034
/gene="gyrB"
/gene="gyrB"
5109..7034
/gene="gyrB"
/EC_number="5.99.1.3"
/note="CDS ID OB0006"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit B"
/protein_id="BAC11962.1"
/db_xref="GI:22775684"
/translation="MSMEDKLTENQVAGDQIOVLEGLVAVRKRPMYIGTSSEKGLH
HLMVATNSIDEALAGYCDHIQVVEEDNSITVDNRGTPVPIQQTGRPALVYIM
TVLHAGKRGGGIVKSGGLHGVASVNALSSLEIVYVHDKGVHLSFKKGVDPGE
IKVIGDITDITGVTHFRPDTEIFETTEYNPDLEQVRLBELAFNKGLSIEDKTD
REQVYHVEGGIISYVEEINRKNKEVLAHPFAEGEDQGISVEVALQVNDGFASNTYSF
ANNIHYEGSGHEVGFSGLTRIINDYAKNGKLIKDGSNISGDDVRGEMTIVSIKH
PDPOEGQTKTKLNSSEVALITDGVSSAFKELYENSTAKITVYKGLMSRARLAA
KKARELTRKSNLEISNLPGLADSSDAIAISELYIEGDSAGSASAGRRHQAI
LPLRKILNVEKARLDRIILSNNEVAMITAGSGVGEFSDISKARVHKIIVMTDADV
GAHIRTLLTFEYRMRPLIEQYIYIAQPPYQVQKQKTVYVANDLEIRILNEIP
KAPKNIQRYKGLGEMNADQLMDTMDPTRTLLQVELSDAIDADQVDMIMGDYEP
RRIFEENQAVYKNDI"
7123..9636
/gene="gyrA"
/gene="gyrA"
7123..9636
/gene="gyrA"
/EC_number="5.99.1.3"
/note="CDS ID OB0007"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit A"
/protein_id="BAC11963.1"
/db_xref="GI:22775685"
/translation="MADINRPVSEINISNEMRTSFLDYAMSVIYARALPDVDMKMP
VHRRLVYAMNDLGMHSDKAYKKSARVCEVIGKHPHDSAVYETMVMQAOTPSRYM
LVHGGNFGSYDGSAAAMARTERMAKISHELLRIDNKOTITDQNDYDTESEVVF
PARPPNLLVNGTSGIAGMATNIPPHHIGITDVLVASIQNDPITIDIMENHLPGPD
PPTAGQILKRGKIRKAYETGSGITIRAKNIEGKNDKETIIVTELPOVYNKALIR
KIAELVNDKVEGITDLDESDRNGMRVILIEGDANAVVNLNLYKHTSLQTFGIN
MLAVDQGPVYLNKQCLSYLIEHOKVYIKRTOPELNKARAHILLEGRIALDHL
EVELIISKTKADTARIGMERFRLSEQOALIIDMLQRLTGLERKTEDEYVLEOA
LIEBKALIDBEKVLEITRELTETREKREKEDRTEIVAGAGFEDEDLIPENIV
ITLTHQGIKRLPASVYRQRRGGIQQMGCTNDDPFEHLVSTSTHOTIILFTNKGIAK
VYKAGYEIPEFSRTAKGIPITINQVKGVMVNAVSNQFDEBAVYLFPTKHGIAR
RTSLEKFNIRKGGILVAGLEDEDLISVRLDQCKEMKATKNGLIRFETQIRSM
GRTAGYVKGISLRGDDVVSMEIIEPSSKILHVNKGGKQTDSEVRIRNGGCVF
TCKLDEKGNVAVKAYNAGDDMLITITAGLIRIQVODISQGTGNTGYHLIRLDGS
EEVAVTRIEKQDEDEVDIEVVEGETTDPSTETNEKE"
9707..10813
/gene="OB0008"
9707..10813
/gene="OB0008"
/note="CDS ID OB0008"
/codon_start=1
/transl_table=11
/product="hypothetical conserved protein"
/protein_id="BAC11964.1"
/db_xref="GI:22775686"
/translation="MKMKTSYLMGCVYLQNDIISKSNPRLAKGTLENDIOIYOK
LIDVIDIDTLVNDKPSKNVNEHSEKREBQDINRSHSVOEKNGPFLKRYAVSAVYKD
QFLWKQOSMPVSIPLDKNALIIPLEQLEQDLAVNTFSLHQFARTEYTLFHSVATSLI
SAFTAKEMGYRGEIMIQVLAGVLCDDGMALDIDVIOKQGPLTSKPKNTKGRPQTS
YVMTKIDPLSLTSGVQLAVLQHHERGNGEYPLGWTLNKVVHYARIIVAVDSYHAMATT

TCYRKARNIVNIEELIAVSTIRLYDDVNTFNMTNIFSNWLENSKVLANNGVGEVVF
 NKEHPRPIVKSNDSDIDIFSLQELPELYIDKILG"
 gene
 complement (10876. .11829)
 /gene="OB0009"
 CDS
 complement (10876. .11829)
 /gene="OB0009"
 /note="CDS ID OB0009"
 /codon_start=1
 /transl_table=11
 /product="hypothetical conserved protein"
 /protein_id="BAC11965.1"
 /db_xref="GI:22775687"
 /translation="METNQLFTHLAENASQFLRTCYONIDEQTKSYONGTFEMHY
 LQHQLOFYKNGESESIVRPVLYFYGMVHLTKMLTKRDPYESTALLAHGVTSRKR
 KRDYOFEDEVKIOOQGLPEYFVRHLYOENPPLKASRLILANIPELYVMNQ
 RETLIDVGRINNKOLVFPSTLIDHVLTKNAPVHRPLRPKEKERETLRIDLO
 NVIMQDGFPPFYMNDQRLYFPTRSDIYTPYSILAHYILVLSMLCYETEMWGL
 MWSKEDRAYPMIEHLEFVTRRKIPYLIEKHLSTYSDCKGK"
 12072. .13541

Alignment Scores:

Prod. No.:	1.16e-08	Length:	300050
Score:	245.50	Matches:	104
Percent Similarity:	38.17%	Conservative:	67
Best Local Similarity:	23.21%	Mismatches:	158
Query Match:	11.01%	Indels:	119
DB:	1	Gaps:	20

US-10-069-062-7 (1-432) x AP004593 (1-300050)

QY 8 ProGlyLysAlaPheLeuAlaGlyGlyLeuValLeuGluProIleTyrAspAlaTyr 27
 DB 248726 CCAGGCAAACTAATGATGCTGAGAGGTCGCTGATTAGAACCTTATCAAAACTTGA 248785
 QY 28 ValThrAlaLeuSerSerArgMetHisAlaValIleThrProIleGlyThrSerLeuLys 47
 DB 248786 GTATATGCGATTATGATACGATGATGCAACCATTAACA-----GAC 248827
 QY 48 GluSerArgIleLysIleSerSerProGlnPheAlaAsnGlyLys-----TrrGluTyr 65
 DB 248828 GCAGAAATTAATACATTTCTCTAGAAAGACTTTGGCTTACATATGCTTGGTCTTAC 248887
 QY 66 HisIleSerSer---AsnThrGluLysProArgGluValGlnSerArgIleAsnProPhe 84
 DB 248888 CAAACGGAAGCTGAATAATAGAACGACGATGAGGATTTCT-----TTT 248935
 QY 85 LeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnProThrGluAlaPheAspLeu 104
 DB 248936 GTGAAGAGCTGATCTACATTAATGATGATGATGATGATGATGATGATGATGATGAT 248983
 QY 105 GluIleIleIleTyrSerAspProGlyTyrHisSerGlnGluAspThrGluThrLysThr 124
 DB 248984 -----ATAGAAATCCCTTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 249031
 QY 125 SerSerAsnGlyLysLysThrPheLeuTyrHisSerArgAlaIleThrGluValGluLys 144
 DB 249032 GCTTCCGGT-----GTGAAA 249046
 QY 145 ThrGlyLeuGlySerSerAlaGlyLeuValSerValAlaAlaThrSerLeuLeuSerHis 164
 DB 249047 TACGGCTCTAGGGTCAAGTGGCGGTATGTAACATCTGTTGTAACGGCAATCTTCTTCTT 249106
 QY 165 PheIleProAsnValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAla 184
 DB 249107 GTTGCCTCCACGTTAC-----GCAACGAAAGAAATTTTTCGTTAGCGGCGTGTCT 249160
 QY 185 HisCysTyrAlaGlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyr--- 203
 DB 249161 CATGCTGTACACAAAGTAT---GGATCTGGTGGCGGATATGTCAGGCTTCACATATGG 249217
 QY 204 GlyLeuIleValTyrArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeu 223
 DB 249218 GGTTTGCACAAATATGATGCTTCAAGCAGATTGGTTATGTAATCATAT----- 249268

QY 224 GluSerAspProGluLysPheProThrGluLeuLysLysLeuIleGlySerAsnTrpGlu 243
 DB 249269 -----AAGAAATGCTTCAACGATATGAAAGTTGATATGAGTGGAG 249313
 QY 244 ---GluLysHisGluLysGlyThrLeuProTyrGlyIleLysLeuLeuMetGlyAspVal 262
 DB 249314 TATTATCATCAAGAAATATTCGATACCTGATTTTCTTAAAGCTGCTGCTGTTGGACA 249373
 QY 263 LysGlyLysSerGluThrProLysLeuValSerArgValLeuGlnTrpLys----- 279
 DB 249374 GGTACTGCTGCATCAACAAAGAACTGTTCACACGATTCGTTATTATAAGTAAGTAT 249433
 QY 280 -----LysGluLysProGluLysSerSerValTyrAsp 291
 DB 249434 TCAGATGCATATGACATTTTCTTACAAAGATATTAAGAAAGCGGCGCTATTTTATGAT 249493
 QY 292 GlnLeuAsnSerAlaAsnLeuGlnPheMetLysGluLeuArgGluMetArgGluLysTyr 311
 DB 249494 GGAATGAAATAGCGAGAACCTTCTATGTTAATGAC----- 249529
 QY 312 AspSerAspProGluThrTyrIleLysGluLeuAspHisSerValGluProLeuThrVal 331
 DB 249529 ----- 249529
 QY 332 AlaIleLysAsnIleArgLysGlyLeuGlnAlaLeuThr-----GlnLysSerGlu 348
 DB 249530 -----GAGTAAAGAAATAAGACAGACATTAGCACAGCTAGCACAAAGCTTAA 249580
 QY 349 ValProIleGluProAspValGlnThrGlnLeuLeuAspArgCysGlnGluIleProGly 368
 DB 249581 ACACCAATTTGAAACATCGATGCTGACAAATATATGACCTTCTGACAGTTAGGTGAT 249640
 QY 369 CysValGlyValValProGlyValAlaGlyGlyTyrAsp---AlaIleAlaValLeuVal 387
 DB 249641 -----GCTGGAAACCATCAGATGCTGCTGAGAGCGGATTTGCTATTTATGATGC 249694
 QY 388 LeuGluAsnGlnValGlyAsnPheLysGlnLysThrLeuGluAsnProAspTyrPheHis 407
 DB 249695 TCTAAAGATCAACGACAGATCTT-----TTTCAT 249724
 QY 408 AsnValTyrTrpValAsp-----LeuGluGlnThrGluGly 420
 DB 249725 -----GCTTGGCGAAGTCCGAATCAAGCCCTTACATTTACAGCTTCATTAGAAAGCT 249778
 QY 421 -----ValLeuGluGluLys 425
 DB 249779 GAGCAAGTTTATATCGAAGAGGAA 249802

RESULT 19
 AX441245 728 bp DNA linear PART 02-JUL-2002
 LOCUS AX441245
 DEFINITION Sequence 4 from Patent EP1209236.
 ACCESSION AX441245
 VERSION AX441245.1 GI:21690241
 KEYWORDS
 SOURCE Gossypium hirsutum (upland cotton)
 ORGANISM Gossypium hirsutum

REFERENCE
 1. Meisner, R. and Lechelt-Kunze, C.
 Eukaryota, Viridiplantae, Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustos II; Malvales; Malvaceae; Malvoideae; Gossypium.

AUTHORS
 TITLE Phosphomevalonate kinases from plants
 JOURNAL Patent: EP 1209236-A 4 29-MAY-2002;
 BAYER AG (DE)

FEATURES
 source
 1..728
 Location/Qualifiers
 /organism="Gossypium hirsutum"
 /mol_type="genomic DNA"
 /db_xref="taxon:3635"

BASE COUNT 213 a 151 c 144 g 219 t 1 others


```

gene
CDS
3365..3368
/gene="ST0848"
3365..3368
/note="PIR:S74037 percent identity: 43.293 in 164aa."
/codon_start=1
/transl_table=11
/product="167aa long conserved hypothetical protein"
/protein_id="BAB5861.1"
/db_xref="GI:15621868"
/translation="MDTSFLIDVWKYDKEDLLFDYTTVFLSESVALNIRTEPLIMI
SEMAKGRILKEETVDRKALVDVDRPGRADYBEAVCVLGGKELMDVYTE
NGVPAKKEIIEYNSVNYRGIDILYISQKGLSDPINEVKYINISKHYSKRIR
EYKGIET"
4482..5090
/gene="ST0849"
4482..5090
/note="PIR:T00006 percent identity: 53.247 in 77aa."
/codon_start=1
/transl_table=11
/product="202aa long hypothetical pepk protein"
/protein_id="BAB5862.1"
/db_xref="GI:15621869"
/translation="WGITKDISKNDIKNAIIEITSGKSTFSKQELLEKTKRKSNT
INILIEDLVKEGFIKKIKQKKNKEIQMFIEGKVDENEGKEDIEKSLDQIAL
ESAFDVLISKYERLIEDLKKIIFRDIRPTEDLDVREFANSLGHASIKDLRLOGL
LSLEEFMYRFYRIILQNYELIPGEEGFIKGVYGIIRKR"
5099..6433
/gene="ST0850"
5099..6433
/note="motif=ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="444aa long hypothetical protein"
/protein_id="BAB5863.1"
/db_xref="GI:15621870"
/translation="WCNPSYTVARYFDNVLVGEWDSARKSLINININDGNVT
IGSPGMGKTTILNAVVDLKNRDFVLYLDSASDKAMNYIRKLEYSK
FYVLSLHKKEIGYKGLYISKDFKMLRHOESSDARIKLCYFYEKIDIGLEFL
NULKTSQLFKVGILIDETSKSEILIDELHINKYKIPAIITLIPQVYKKNALVR
RLTENQIEKLSDKKEELIKAYCEKYELLISIDVRAETLNALDPAKRYEAL
NECKKGAYYIDECIKDYLOKTPIDESLDDYSTLEKRIKLEIKQAAPVIKYI
HERGKGLDEYGVTDIYFETENTYIGDVKLSNEKLNKIKVKKLENYQIKDKKV
KEIVTVNNVNLSDFTIIPVNNLSIKKIVDNDKEKRDILVGMNRLGIV"
6481..6894
/gene="ST0851"
6481..6894
/note="Swiss_Prot:Q10858 percent identity: 28.289 in
304aa."
/codon_start=1
/transl_table=11
/product="425aa long hypothetical transporter"
/protein_id="BAB5865.1"
/db_xref="GI:15621872"
/translation="MEKKTLISQALALGIGNIGAGIFVMAGVSTIAGPALAFL

```

```

ITAVYAMSGVGNNAELASVPFVEGVYSPALLSIGETIGLVGMFVIGVAINSGAT
ALGFSGYITTEFSGFLYFLAIIIVLIIIDVLGKLAIESILVNLGLV
PSITLVISGRIQNSPFLPHGIVGLVNASIAPAYSGFNTIATLPVYGEERTV
PKATVLSLISLNFELMUTVFSVNDIYGNKRNTPAPQVLSITHAPFYFVNL
ATLATTYVLSITARTWOMELNDLPPKRGKSLILIGVMTASJAGLVBV
IALVSNFGIIVPSYMSGIEVAVVRHGRGKFSKFPFVQIISVLSGVFLSLGPO
SLILGFVTLILGLVHSIHKEIMKKEGIFSRKK"
8837..9703
/gene="ST0853"
8837..9703
/note="PIR:C75043 percent identity: 28.821 in 229aa."
/codon_start=1
/transl_table=11
/product="288aa long hypothetical protein"
/protein_id="BAB5866.1"
/db_xref="GI:15621873"
/translation="MILGKLAELKEGKIYKIYGYEKPSLQNIATVKNYGLK
REISGRIIEIGISLYKGGDEKVPFATGGEKIPYLPYPLVIDMSLFELDE
BEKKTLOVNLTLHVRKFLMDGNALINSPISEKARNIDNETNCLVDPYGD
IIANEELIRBDVTFIIGIVDKGRRLKYATSKAEKVCYKTIIRGSIYGVDEIN
KIADIIIVKRGKELEKAIITQSSNDKIALIVDVRGELIEEVRMANAKVY
KILIRLGVNAS"
complement(9806..10966)
/gene="ST0854"
complement(9806..10966)
/note="Swiss_Prot:P19780 percent identity: 35.897 in
117aa."
/codon_start=1
/transl_table=11
/product="386aa long hypothetical transposase"
/protein_id="BAB5867.1"
/db_xref="GI:15621874"
/translation="WTKDPVSAPAIIDISQSLTRAKSELVYKQKPSVTYKVEFP
YNNGIEELIKFLBGVNGILFATGVYVYVYHEKLTGRNLTVPVNPALHLELTKT
DKLDAORILVAVMTGVIRGYSYIPGIEIKELBELTGRNLTVPVNPALHLELTKT
GKYEIPFDPKROSLKELKAGSELSEKELKEGLGNDLDAELTLKQVLELKS
ENMVEKEDWAIISKIPKVEIISKIPGIGLTAATVTFEFGISFSPSKARAVAGF
APRTQSGNSHSGMIRGNKRIIRGFPYLARSAWKEPPEVRRLIRGSGVQAT
CALAGKACIYHVILKDGVIYKIKRLIRPGSEVNVKPDVGDALDSLSF"
10218..10748
/gene="ST0855"
10218..10748
/note="PIR:F71169 percent identity: 37.313 in 67aa."

```

Alignment Scores:

Pred. No.:	2.94e-07	Length:	253050
Score:	228.00	Matches:	98
Percent Similarity:	38.94%	Conservative:	71
Best Local Similarity:	22.58%	Mismatches:	121
Query Match:	10.22%	Indels:	144
DB:	1	Gaps:	18

US-10-069-062-7 (1-432) x AP000984 (1-253050)

```

QY 6 SerAlaProGlyLysAlaPheLeuAlaGlyTYrLeuValLeuGluProIleTyRhp 25
DB 135243 TCTGCTCCGGGAAATATTAATGATTCGAGTTATTCAGTA-----GTATTTGGC 135193
QY 26 -----AlaTyValThrAlaLeuSerSerTgmet----- 35
DB 135192 GGGATTTACACGTTATAGCATTAATAAAGAGAAAGATGCGACATATAAAAGTTCTAAC 135133
QY 36 -----HisAlaValIleThrProLysGlyThrSerLeuLysGlu 48
DB 135132 AATTCAATTTTGAACAACAATACGTACTTCAAAAGATTAAGAAAGAAAGATTAATCGAA 135073
QY 49 SerArgIleLysIleSerSerProGlnPheAlaengIyluTgpluTyRhIleSer 68
DB 135072 AGTGTAAATTCAGTTTAAAGAAAGATTCGAAAC----- 135037

```

QY 69 SerAenthGluLysProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThr 88
 Db 135036 -----TACCCCCCATC----- 135025
 QY 89 IlePheIleValLeuAlaTyrIleGlnProThrGluAlaPheAspLeuIleIleIle 108
 Db 135024 -----CATGGAGATTA 135013
 QY 109 TyrSerAspProGlyTyrHisSerGlnGluAspThrGluThrLysThrSerSerAsnGly 128
 Db 135012 TTTAAGCAT----- 135004
 QY 129 GlnLysThrPheLeuTyrHisSerArgAlaIleThrGluValGlnLysThrGlyLeuGly 148
 Db 135003 ---AAGATTTTCAGATCATGTTAAA-----AAACTGGTTTAGA 134965
 QY 149 SerSerAlaGlyLeuValSerValAlaIleThrSerLeuLeuSerHisPheIleProAla 168
 Db 134964 AGTTCTTCAGCATCAACTGATGCTGCTGTTGATTATTTATTTATTTCAAGAAAT 134905
 QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
 Db 134904 TTA-----AACAAAGATGAAATCTTAATTAATTCACAGAAAGCAATTAATTAGG 134854
 QY 189 GlnLysLysIleGlySerGlyPheAspValAlaIleThrAlaIleTyrGlyLeuIleValTyr 208
 Db 134853 CAAAAAGTATGGGAGCGGTTTGTATAGCTCAGCTTTATAGTTTCAATAGTTTAT 134794
 QY 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGlnLysSerAspProGlu 228
 Db 134793 AGAAGATTTTAT-----GATATAGAG 134773
 QY 229 LysPheProThrGluLeuLysLysLeuIleGlnSerAsnProGlnGluLysHisGluArg 248
 Db 134772 AAGATGATAGTGTATTAACCCCTTAAGATAGCAACTATAGAA----- 134728
 QY 249 CysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValIleGlySerGlyThr 268
 Db 134727 -----ATGCTCTTAAGTTTCATAGCGTAAGCTTCAAGTACT 134692
 QY 269 ProLysLeuValSerArgValLeuGlnIleTyrLysGlnLysProGlnLysSerSerVal 288
 Db 134691 GTAACTCTGATGCTAACTTTATTGAA-----AAAGCAATTAACGAGGATTAAAGAA 134638
 QY 289 ValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGlnLeuArgGlnMetArg 308
 Db 134637 GTAAATGAATACATTGACAGAGAAACATTAATGCAATTAATAA----- 134587
 QY 309 GlnLysTyrAspSerAspProGluThrTyrIleLysGlnLeuAspHisSerValGluPro 328
 Db 134586 -----TTAGTAAATTTAGAGAACCCATAGAAC 134557
 QY 329 LeuThrValAlaIleLysAsnIleArgLysGlyLeuGlnAlaLeuThrGlnLys---Ser 347
 Db 134556 GTGAAACTGCT-----AGAAATTTCTAAATGCTTTAGCTTAAATAAATAGTT 134509
 QY 348 GluValProIleGluProAspValGlnThrGlnLeuAspArgCysGlnGlnIlePro 367
 Db 134508 GGTGTAAGATGAAATGAGAAAGATAGAAAGATTGATGAATGAGCTGAAAT----- 134455
 QY 368 GlyCysValGlyValValProGlyAlaGlyTyrAspAlaIleAlaValLeuVal 387
 Db 134454 GACGATTAATTCCTTATCCCTGAGCAGAGAGAGAA-----TCACTATTGTGT 134404
 QY 388 LeuGlnAsnGlnValGlyAsnPheLysGlnLysThrLeuGlnAsnProAspTyrPheHis 407
 Db 134403 TTAGGTAAAGATTTCAAAGGTGAAAGAAAGTGGGAAAA-----GAA 134359
 QY 408 AsnValTyrTyrValAspLeuGlnGlnGlnThrGlnGlnVal 421
 Db 134358 AACGTTATAGTTATAGATTAAAGAGAT---GAGGGTTTA 134320
 RESULT 21

AEO06889
 LOCUS AEO06889 9921 bp DNA linear BCT 05-JUL-2001
 DEFINITION Sulfolobus solfataricus section 248 of 272 of the complete genome.
 ACCESSION AEO06889 AEO06641
 VERSION AEO06889.1 GI:13816365
 KEYWORDS
 SOURCE
 ORGANISM
 Sulfolobus solfataricus
 Sulfolobus solfataricus
 Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 Sulfolobus.
 1 (bases 1 to 9921)
 AUTHORS
 She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G.,
 Ameyez, M.J., Chan-Weiher, C.C.-Y., Clausen, I.G., Curtis, B.A., De
 Moers, A., Erasmo, G., Fletcher, C., Gordon, P.M., Heikamp-de
 Jong, I., Jeffries, A.C., Kozera, C.J., Medina, N., Peng, X.,
 Thi-Ngoc, H.P., Redder, P., Schenk, M.B., Theriault, C., Tolstrup, N.,
 Charlebois, R.L., Doolittle, W.F., Duguet, M., Gaasterland, T.,
 Ragan, M.A., Senses, C.W. and Van Der Oost, J.
 The complete genome of the crenarchaeon Sulfolobus solfataricus P2
 Proc. Natl. Acad. Sci. U.S.A. 98 (14), 7835-7840 (2001)
 21332296
 MEDLINE
 PUBMED
 11427726
 REFERENCES
 2 (bases 1 to 9921)
 She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G.,
 Ameyez, M.J., Chan-Weiher, C.C.-Y., Clausen, I.G., Curtis, B.A., De
 Moers, A., Erasmo, G., Fletcher, C., Gordon, P.M., Heikamp-de
 Jong, I., Jeffries, A.C., Kozera, C.J., Medina, N., Peng, X.,
 Thi-Ngoc, H.P., Redder, P., Schenk, M.B., Theriault, C., Tolstrup, N.,
 Charlebois, R.L., Doolittle, W.F., Duguet, M., Gaasterland, T.,
 Ragan, M.A., Senses, C.W. and Van der Oost, J.
 Direct Submission
 Submitted (24-Apr-2001) Europe/Canada joint project: Copenhagen
 University, Denmark; Universite Paris-Sud, France; Wageningen
 University, The Netherlands; Institute for Marine BioSciences &
 University of Ottawa, Canada
 location/Qualifiers
 1..9921
 /organism="Sulfolobus solfataricus"
 /mol_type="genomic DNA"
 /db_xref="taxon:2287"
 /complement(101..1012)
 /gene="SSO2981"
 /complement(101..1012)
 /gene="SSO2981"
 /EC_number="5.1.3.2"
 /note="Energy Metabolism, Sugar metabolism"
 /codon_start=1
 /transl_table=1
 /product="Epimerase, putative"
 /protein_id="AAK3085.1"
 /db_xref="GI:13816365"
 /translation="MKILITPGCTGPGVSGNFFRYAISKSHDVLVYARSKNDYTKALONI
 GANTVFSYDNTLKIDICLVYFIGAMWTQDPDEFYLOVFPTEIGRKFVYNSGKFPV
 ISSIGVSENIIESKRPPIAESPHBEGKPNTHGITYLKEGRNISKPNVYILRPYIV
 YGYSRIILMRAFMVLVSHGIGLKDNNFSVSTRNSKATLELACKYRSYFYITDK
 EPAVLTSIFSDAVKAINREISMLPFSYKLEIPKXSHMKMAYDVLRSRLVYSYK
 AARLANTYDENVRLIEFKEMAKYKII"
 1093..1305
 /gene="SSO11614"
 1093..1305
 /gene="SSO11614"
 /transl_table=1
 /codon_start=1
 /product="Hypothetical protein"
 /protein_id="AAK3086.1"
 /db_xref="GI:13816367"
 /translation="MNYSKFARFKEMALTTNDEVILPHKLRKIVEIKRNPDIIVLR
 LAGLIDPDALVALYLRNSYKIVET"
 1351..2313
 /gene="SSO2982"
 1351..2313
 /gene="SSO2982"
 /codon_start=1

```
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK43087.1"
/db_xref="GI:13816368"
/translation="MKRIKIEPFWGVNIPPIFLAMITVMAIGSLSPLELPHPYFMYL
GAVYAFEGMIQRLEFPAPKYNLAHLASLILALPIQYFOIASVVLITTEVVALKOLK
MYGVNFKLPIKALVILVSSPSSIIIAVFNWVLLITPILLYLIGVNVFVSYNLKT
PVEGHQLPIFLIILSYFPIIPPIGVVFLITRYKTFPKSIGISLSLIVIP
LILVYRGDFVAAFTIGLMSNLFPSCITYSTSRNNKVVISILSDPLAILNPFYEI
SGIFWVAVITPYILIKONRYLITSLGSMRIKRNKRGSP"
/complement (2276. .3505)
/gene="SSO2983"
/complement (2276. .3505)
/codon_start=1
/transl_table=11
/product="Conserved hypothetical protein"
/protein_id="AAK43088.1"
/db_xref="GI:13816369"
/translation="MRKRVLAIGVNFPSGVSVHINVLKELVKKMGDEVFLVPSFL
VRNRYEVLKDLERNAHISVHSVYETKEAEVGLPYRISHVTYLRMNNLGEDRKYID
ELKDIKPDVITDMHEDSITLRSLSYHIAKRLNKLVLHDEPPRNSFGGKRYKGLIK
GLAYDTLMWFYKFDKRAFLRSINDGILGIGVSAFPYLSGIEIASHVRLKXVEV
GNAFNKDIIYKRTKRGKEDYAVFARLVPQKIKELPKIADLDSKIVFQKIFEN
DSILMSHPKIEYKGRPIEBVNTVAKAKVLIPSHODGFSLVVLDITLALGTSVAY
DIPIRFEVSSLSKPVRTVREYDIKEMASTANSILRMKDEEYSEHDEKVKKFLHLS
SNANVAKETHDPLSAPL"
/3693. .4211
/gene="SSO2984"
/3693. .4211
/gene="SSO2984"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK43089.1"
/db_xref="GI:13816370"
/translation="MDPFLISLILGLSHGIEPDHVAFLRLSKRWKIVOFALSHSNGF
VILAIPLVILIGNKKLEIISNIGIIFSLILLOIAPKREIDIGNKKGLQGAFAVI
TPKXVLIVIASPAVSILYSIEVISVFIASANSISLSLNVPRVKKYKIVDVGAL
LMTYILIFILIN"
/complement (4442. .5407)
/gene="SSO2985"
/complement (4442. .5407)
/gene="SSO2985"
/codon_start=1
/transl_table=11
/product="Transposase ISCI439"
/protein_id="AAK43090.1"
/db_xref="GI:13816371"
/translation="MOTMILPKTELKSLAITLATNNINVTISODLDEIYKAAPSLLTG
NGKYYLKVVRREGKVISKQRTFKFYPIYREKGEINVADEGTGLTGEKOEKAE
GFLVNMKRGKGVMSRLDLYPLRLPLVEVDLRSDSPSOILASREVRSOYMEIDY
VVAADAGFLNLGVTKEMPVTIYRGSKNLGFPELSVPIVEKRYEKDKVYVAYRTE
FEGLYYDVAVYKGRPHFMFVTFEGDYEELAEELRLMWVEEGKVKARIRYRK
LSNKIFLPLYYTVLDSANLVNHLFNFSTCKVLSFDSFVYLL"
/complement (5630. .5932)
/gene="SSO2986"
/complement (5630. .5932)
/gene="SSO2986"
/codon_start=1
/transl_table=11
/product="Conserved hypothetical protein"
/protein_id="AAK43091.1"
/db_xref="GI:13816372"
/translation="MRTRDDIISDILEALIDSKINRISIMKNVNISASLAKKYSML
AEBGLIQDNGEKVLEKGRKILIEQLRNNRKWELKELANIYEVREKPVVHFQFSK"
/complement (6413. .7021)
/gene="SSO2987"
/complement (6413. .7021)
/gene="SSO2987"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
```

```
/protein_id="AAK43092.1"
/db_xref="GI:13816373"
/translation="MEFLSKRERVLVAIAQPSGTYTSOLSAIGYILSRNTIOETI
ENLYKGYNTNLKQDEIRYASKEVRNAILNLEPHKRLAKNLSLKKYVDISKME
KQQAQVAFRGVAKDLSSISLSLDDMPALITPTEVLENNIYREFLSKLSPLMQ
ESMTDEINAFSLISKYRGEKEAEIKSTYLOKIGABSO5"
/7073. .8044
/gene="SSO2988"
/7073. .8044
/gene="SSO2988"
/codon_start=1
/transl_table=11
/product="Phospho mevalonate kinase, putative"
/protein_id="AAK43093.1"
/db_xref="GI:13816374"
/translation="WIKVSAPCKIIMTSYSYVEGCIISHVIAVNRKVSCLREIKED
SLIFHTSYGHEKNSGNELNSVLDTRFRLSQLPGYEIDLNDPEFIIIDGKTGLGS
SSAATVSLTACLKYAIGHKLDLFEIHKLAQIANRYRQKIGSGPDIASAVFSIYKR
FTDLDMDFYFEKMLGNVDMMLGFTGSSSETVGVRKFEKSNLDFEKEIWRLLDEE
NYMALIKLILKMLDEAVEHIKLGRKYLVYIAERYVGRVLSKMESELKIAIEEGALV
ALSPPAGGDSIFALGNDINRYREMSKRGIFIIDVKDEGLRESLN"
/complement (8049. .9026)
/gene="SSO2989"
/complement (8049. .9026)
/gene="SSO2989"
/3693. .4211
/codon_start=1
/transl_table=11
/product="Diphosphomevalonate decarboxylase, putative"
/protein_id="AAK43094.1"
/db_xref="GI:13816375"
/translation="MLKSVTVSAPSNIAVVKYKRGGRERLNLPLNLSITLDOLS
VITKTLNDKNIIVYNDRIISGDEMEKRAVRGLDTPKTIQGEHPHYKESKRPINA
GLASSAAGIAPALAEFLNLELNLSELSKIRLGSQACSMFGVYVNRKGERD
GEDSYCYOIFRHDVSELVDIIPILSEKIKISSKGMIRKSTSELMECRVLEKT
FNEVEIARNRBEKKFVYLMKMSMAVILDSWSPFFYINDTSIRIMEMWIDYGA
GTYPPAGNPHIFFTERNIGDLEFLSKLEIKRIIVSKVGDPKLSRB"
/complement (9045. .9728)
/gene="SSO2990"
/complement (9045. .9728)
/gene="SSO2990"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK43095.1"
/db_xref="GI:13816376"
/translation="MEKVIDKYIIRTDOSDNVINDLVKILRKTNIKAYNKEFELH
NKVSIRAIRNRIILNLSNLYIKDMEDILSESELVYTFRFGIEFNHIPSRRLLDLKLEA
TGLPYSKVDVEFDVYRWTFINGFTFIDKSLSEATYLSLILEKVLPEFNLGRIRKX
DWRALLLKYYGIRDLDLLEKLDLGLRIENDNEIILIDNISISKKIGIFPKGNVSKKE
LYEIVKAKN"
/BASE COUNT 3006 a 1907 c 1503 g 3505 t
ORIGIN
```

Alignment Scores:

Pred. No.:	7,686-09	Length:	9921
Score:	225.00	Matches:	99
Percent Similarity:	40.00%	Conservative:	69
Best Local Similarity:	23.57%	Mismatches:	142
Query Match:	10.09%	Indels:	110
DB:	1	Gaps:	16

US-10-069-062-7 (1-432) x AE006889 (1-9921)

QY 6 SerAlaProGlyLYaLaAphelaUaGlyGlyTYrLeuValleuGluProIleTYrAsp 25

DB 7085 AGCGACCTGGCAAAATACTTGGATGAGGATTACAGGTAGCT----- 7129

QY 26 AlaTYrValThrAlaLeuSerSerayGethGalaValIleThrProLYaGlyTYrSer 45

DB 7130 -----TTTGGAGGACATATCATCATGTAAATAGACGTTAATTAAGAGTGA 7174

QY 46 LeuylSerArglIleLysIleSerSerProGlnPheAlaSerGluTrpGluTrp 65
 DB 7175 TGTAGTTTGAAGAATTAAGAGAGATGTTTATTTTCCATCTAGTATGAGGCAC 7234
 QY 66 HisIleSerSerAsnThrGluLysProAlaGluValGlnSerArgIleAsnProPheLeu 85
 DB 7235 TTCAAAATTCGTGTAATGATTA-----ATTAAATTCAGTCTTCAATCTCTTTTACA 7285
 QY 86 GluAlaThrIlePheIleValLeuAlaTrpIleGlnProThrGluAlaPheAsnLeuGln 105
 DB 7286 GAAAGCTTTCACAACTTCTCAGGTTAT-----GAA 7318
 QY 106 IleIleIleTrSerSerProGluTrpYrHisSerGlnLysAsnProThrLysThrSer 125
 DB 7319 ATAGATCTATATACAT----- 7336
 QY 126 SerAsnGluLysThrPheLeuTrpYrHisSerArgAlaIleThrGluValGluTrp 145
 DB 7337 -----AAGAGTTT-----ATAATAGATGTATAAAAAAGC 7366
 QY 146 GlyLeuGlySerSerAlaGlyLeuValSerValAlaIleThrSerLeuSerHisPhe 165
 DB 7367 GGAATAGTAGTCTTCCGCT---GCTACAGTTTCCCTTACCGCTTCTGTATTAAGCT 7423
 QY 166 IleProAsnValIleSerThrAsnLysAsnIleLeuHisAsnValAlaGlnIleAlaHis 185
 DB 7424 ATCCACGGCAACTGATTTGTTTGA-----ATACATAAATGTCACAAATTCGCACAC 7477
 QY 186 CysTrpAlaGlnLysIleGlySerGlyPheAsnValAlaThrAlaIleTrpGluTrp 205
 DB 7478 TATAAGAGCAAAAGGATTTGTTGATGCTTGTACATGCTCAGCTGTTTGGTATGC 7537
 QY 206 IleValIleTrpArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGlnSer 225
 DB 7538 ATAGTTTCAAGAGATT-----ACT 7558
 QY 226 AspProGluLysPheProThrGluLeuLysIleLeuGlnSerAsnTrpGluLys 245
 DB 7559 GATTAGATAGATGAGATTTGATTTGCAAAAACCAATTTGGAAATTAGCAT----- 7612
 QY 246 HisGluArgCysThrLeuProTrpGlyIleLysLeuMetGlyAspValIleGlyLys 265
 DB 7613 -----ATGATCTCTGATTTTACCGGAAAGAGT 7639
 QY 266 SerGluThrProLysLeuValSerArgValLeuGlnTrpLysGluLysProGluGln 285
 DB 7640 TCTGAACCGGTGTTGTTGATGAAATTTGAGG----- 7675
 QY 286 SerSerValValTrpAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGluLeuArg 305
 DB 7676 -----AGAGTAATTTGATGATTTTAAAGCAAAATATAG 7708
 QY 306 GluMetArgGluLysTrpSerAspProGluTrpYrIleLysGluLeuAsnHisSer 325
 DB 7709 AGGCTT-----ATAGATGAGAGAAATTTATATGCAATCAAACTCATATAA 7753
 QY 326 ValGluProLeuThrValAlaIleLysAsnIleArgLysGlyLeuGlnAlaLeuThrGln 345
 DB 7754 TTGAATTAAGCTTGAAGAGCTGTGAGACATTAAGTTAGGAAGAGATTTGATTAAT 7813
 QY 346 LysSerGlu-----ValProIleGluProAspValGlnThrGlnLeuLeuAsp 361
 DB 7814 ATAGCTGAGCGTATTTGTTGTTTAACTGTAAGTAATAAATGAGAGAGAGTTGATATAA 7873
 QY 362 ArgCysGlnGluIleProGlyCysValGlyValValProGlyAlaGlyCysTrpAsp 381
 DB 7874 ATAGCAAGAGAG-----GAAAGTCCGTTAGTACCTTATCTCCGCGGCGAGAGAT 7930
 QY 382 AlaIleAlaValLeuValLeuGlnLysGlnValGlyAsnPheLysGlnLysThrLeuGln 401
 DB 7931 TCAATT-----TTTGCATTAGCGCAATGATTTAAATAGGGTAAAGGAGCGTGAAGTAA 7984
 QY 402 AsnProAspTrpTrpPheHisAsnValTrpTrpValAsnLeuGluGlnThrGluVal 421

Db 7985 -----AGAGTATTTCAATTAATGATGAAAGAGAC---GAGGATTA 8026
 RESULT 22
 AL935257/c 302050 bp DNA linear BCT 21-FEB-2003
 LOCUS Lactobacillus plantarum strain WCFS1 complete genome; segment 6/11.
 DEFINITION
 AL935257 AL935263
 ACCSSION
 AL935257.1 GI:28271236
 VERSION
 AL935257.1
 KEYWORDS
 Lactobacillus plantarum WCFS1
 Lactobacillus plantarum WCFS1
 Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
 Lactobacillus.
 REFERENCE
 1 Kleerebezem, M., Boekhorst, J., van Kranenburg, R., Molenaar, D.,
 Kuipers, O. P., Leer, R., Tarchini, R., Peters, S. A., Sandbrink, H. M.,
 Fiers, M. W. E. J., Stiekema, W., Lanthorst, R. M. K., Bron, P. A.,
 Hofer, S. M., Groot, M. N. N., Kerkhoven, R., de Vries, M., Ursing, B., de
 Vos, W. M., and Siezen, R. J.
 Complete genome sequence of Lactobacillus plantarum WCFS1
 Proc. Natl. Acad. Sci. U.S.A. 100 (4), 1990-1995 (2003)
 2 (bases 1 to 302050)
 Kleerebezem, M., and Siezen, R. J.
 Direct Submission
 Submitted (14-OCT-2002) Wageningen Centre for Food Sciences, P. O.
 Box 557, 6700 AN Wageningen, The Netherlands
 Location/Qualifiers
 1. 302050
 /organism="Lactobacillus plantarum WCFS1"
 /mol_type="genomic DNA"
 /strain="WCFS1"
 /db_xref="taxon:220668"
 /note="Single colony isolate of strain NCIMB8826"
 82. 999
 /locus_tag="lp_1715"
 82. 999
 /locus_tag="lp_1715"
 /locus_tag="lp_1715"
 /EC_number="2.5.1.-"
 /codon_start=1
 /transl_table=1
 /product="premyltransferase"
 /protein_id="CAD64141.1"
 /db_xref="GI:28271237"
 /db_xref="GI:28271237"
 /translation="MKKMLTPVFFYLTETIYTAPLNIMFVLGSAIAQHLGTVNL
 NVTGLVSVFIFDLAVVSDNYDYDRAHREAGQKTNPIGRQLDLPVAGVRLALGL
 YLSTLIGIMLVKRTGMLVVLGLVGVIGIFPYAGCPYNATPCEVVALAATYFLI
 ELCTVSVTSGOHLTWSIVGTFFLCLPLTLFFCQLANTRADRDIEDILNRYTTLA
 YLGGKPSVRLIOAFIVIGIWIPLVNVTLGLAPVITITWVLLPIIMQGRPFREYOD
 KKRPITTVASASLFIILYEVFLVLTGTL"
 1098. 1874
 /locus_tag="lp_1716"
 1098. 1874
 /locus_tag="lp_1716"
 /locus_tag="lp_1716"
 /codon_start=1
 /transl_table=1
 /product="potassium transport protein"
 /protein_id="CAD64142.1"
 /db_xref="GI:28271238"
 /translation="METERQOORIAVYHLIVTLTMSAVNVLLSFHIGDMVETR
 ISNGILVPAVDYVRLIRSDRRTPIIHAFOIMGILPMPVVALPRLIGLAMYV
 HHLFWKGLDGKTHDHRITYSNGPIYLSISISATIVLSALPFSVEHOSLSLWV
 ALTRATTVGGDDPHTAVGVAVAGMFGIGTGLTITDPTFOASQNVITIE
 PÖHDELRÖLLIKIDTLITTVYDHLÖNVKRLERSGYKKÖH"
 1967. 2266
 /locus_tag="lp_1717"
 1967. 2266
 /locus_tag="lp_1717"
 /locus_tag="lp_1717"
 /codon_start=1
 /transl_table=1
 /product="unknown"
 /protein_id="CAD64143.1"

```

gene
CDS
/db_xref="GI:28271239"
/translation="MCOAIRRWMLPAPITVROSTFKIHVGLDRHPVPLRAQWYOF
KOSRTLCISVPTTQYAAVHTLAQAOLKAGAGMLPAPEQVPMRTGTTTFPM"
2406..2678
/locus_tag="1p_1718"
/locus_tag="1p_1718"
/locus_tag="1p_1718"
/codon_start=1
/translation="lipoprotein precursor (putative)"
/protein_id="CAD64144.1"
/db_xref="GI:28271240"
/translation="MKTRYQLQVXKDGIRSLTRCRVLAIIIASFTGMLTGCBLAQATT
VPTMQTYHPMSVGRGLNLSRLVATVYGLVCMGLATYLPKPKAK"
3116..4465
/locus_tag="1p_1721"
/locus_tag="1p_1721"
/locus_tag="1p_1721"
/BC_number="2.6.1.-"
/codon_start=1
/translation="amino transferase"
/protein_id="CAD64145.1"
/db_xref="GI:28271241"
/translation="MISQNRDLDOOLIEREDHYMTAARINYYDLVIDHAGALLTDV
DNOYIDLASASAINVGHTRPVVAIOEAKLIHTPAYFHHQPEQRLAERLAKS
ARQIDNEVVGNSGDANDAIIFKFAATNROYIVATDAGSTGYSMSLGSVLSM
VRKMGPLPGIVHPYPCYRILPHETHEILARFEARPEPDSYLPABEPAVITE
PIQDGGIRKAPAEVYQVYDRCHGIIIPADEVNOGMRTRKMSIONFPAIRPDL
MSVGSLASGLPLSAVIGREVEVESIAAPHTFTIANPVCCAAALATIDVLEDELV
ASANGYRAKEQFALQORHPKIGQVRYNGVGLDELTRQSDPDPEKSDVYA
AERGVYMTLKNIIRPQPLVITKTQDLTALVIDEAMAAEQGLVRPQKGI"
4482..5981
/locus_tag="1p_1722"
/locus_tag="1p_1722"
/locus_tag="1p_1722"
/codon_start=1
/translation="amino acid transport protein"
/protein_id="CAD64146.1"
/db_xref="GI:28271242"
/translation="MQLRLRLTKEDPSIYEDKSHLARVLTVKDFALGVGTIVST
SIFTLPGVVAANPAVAVESFVAAVVAGLVAFAVAEMAAMPFAGSAYSWINWFG
EFFGMJAGWALAEYFIALAFVSGLSANLRLGLSLPLALPALSNTGEGGVVDL
IYLVATVLSLSRGISKASRVENLVVYLKIVATPFIYVGTATIHQVNYEYFIVY
HINADGAFRGMOGIVAGVSMIFAYIGFDSIAANBAKMPKTMPPRIGLISIVAY
VLPVAVATLVGKFSYANNAEPVGMALRQAGHYASVIOATVGLMFTALIGMT
LNGSRLYISFGDGMLPKWLGLHRVQNNALLVITVAIIIGAFCEPFAFLQQLSAG
TLIAEFVSLGIYALRRRGVDIAPAFKMPYVPAIAGLFEVNGLDIOAKLY
AGVWFIYGLVIFSYGMRSHYLAERKLSLTDPAKSAKSVADSTVEERD"
complement(6161..6970)
/locus_tag="1p_1723"
/locus_tag="1p_1723"
/locus_tag="1p_1723"
/translation="HAD superfamily, Cof family"
/protein_id="CAD64147.1"
/db_xref="GI:28271243"
/translation="MERKJALDLGGTTNABSKNSATIQTMORIQAAGHIVSITTG
RSYQSOQAIYDQLRLKSPMINENGSLGHLPHQWASEYQFTINKALVIALVYREQG
INEIYLAENKSTWTAGODLIAGQFPPEIDSPILPDTLRPNLSVILLVQDLRRPAI
IMITRPGDOVDVGGPNALLETISGKVKAKGAVIADYFOISRNNITAFGDEH
NLEMLAVAGLGVANANGTDKIATADVTATANNQNGVARYLINDFALAE"
7118..8485
/locus_tag="1p_1724"
/locus_tag="1p_1724"
/locus_tag="1p_1724"
/locus_tag="1p_1724"
/locus_tag="1p_1724"
/codon_start=1
/translation="RNA methylase (putative)"
/protein_id="CAD64148.1"
/db_xref="GI:28271244"

```

```

gene
CDS
/translation="MQLPAEFLQYORLLGDAQAPAFIAFDGPEVEKGRINPKTVTA
TKRATMTAPVPSYSGKXGNALNLEHLAGVVSQBSAMTVEFAPBEPEDL
CAAPGKTHLISYHOGTLVTNINORVATVAGSDNREYRGARTVITNDPDLAL
ELRGEFRLIYDAPCSGGMERKCHDAVOYTPDPAACARORPILTEAYMLAPG
HLIYSTCPABEEDQWAMLLKTYPELEBELATKAGVMDKPEMADGNPELAAAR
LPHLMRGEHFIALVYHGTGVKDKKVEEQNATQRQMTDPLKQANAGLPOLV
LOAHGDOLGVAPAMPVTRKLVFRRPGLIGTFKKKREPEPSYALALASDDSLPTEI
TREOMALVYHGETFELTSAPVAGFRVLTCGGLPVQFGKIYAKTVANFPKGLRPLATS
ENATL"
8544..9526
/locus_tag="1p_1726"
/locus_tag="1p_1726"
/locus_tag="1p_1726"
/locus_tag="1p_1726"
/locus_tag="1p_1726"
/codon_start=1
/translation="unknown"
/product="unknown"
/protein_id="CAD64149.1"
/db_xref="GI:28271245"
/translation="MYCPCGKNDIDALFCENGARLLIPASSAQPTPASVSSAQ
STSSTAASEAPSASTSISAAPPAPEGRSAQHAQONQOPSTSLSWIILVILVA
LCGGYVMEHORQNSTATTAKSSSRATSEKSSSSSSKASASTATOTLMTSTKS
SEIASFMSMODTMSYEGTYDQGSVTIGELNLPDIIKNNOYODKITVNGDTVLKY
TTADTDKQYVVAATDPNASSGVITTYLYFPHNSPVLVSQDSATATTFRFTSSON
TDLQGFAXIANAD"
10021..11367
/locus_tag="1p_1729"
/locus_tag="1p_1729"
/locus_tag="1p_1729"
/locus_tag="1p_1729"
/locus_tag="1p_1729"
/codon_start=1
/translation="sugar transport protein (putative)"
/protein_id="CAD64150.1"
/db_xref="GI:28271246"
/translation="MKERTSKTAGIPIKLSNKIMMSISGFIGVOMAFLLQSSONRIF
OTLQDNPNNHCFILPLAGIYVOPITIGISDRMARLGRRLPYILGTVAIIV
MLLPNSGPFGRVIAALVFGATVAFVATLKSIVNAQPMFMGVNDVNDQSIYAY
GIOSFLNTGVIAIIFPLLTLFQVNTAFSGVIPNSVIFVYGAAILVSTIITI
ARVREYDPTVYAKYIGIDESANKSGNMLALKKAPKFMVLVAVQFCMNAFOYLMT
YSTGMANVHTTNASAGVOAGNWTGILSAVOSVAVAVWSVYLAIVPNNRHLGY
ALSTILGAIGRPSIFRHSOSALVNSPILVGMANVAMNTYPLWTVTNALSGEHMGTYL
GLFNGSTLQIOWASLSLFPALYPLFGNSQANNFIPAGIYLAAGACTVGLIKETYN"
11381..11382
/locus_tag="1p_1730"
/locus_tag="1p_1730"
/locus_tag="1p_1730"
/locus_tag="1p_1730"
/locus_tag="1p_1730"
/codon_start=1
/translation="maltose phosphorylase"
/protein_id="CAD64151.1"
/db_xref="GI:28271247"
/translation="MKRIFDVNPWHVLTTRDPENKRIQBSNTSLGANGMGRGFEF
DYSGDPLGIVGVWYDPTKRVGMWKNQYBEFKVINANFIKINKLNGTKIDLA
TANSEDFLIDMDHGLTRSFIVDQDQVRVTERELSYAKSLSVQKTYFENLSD
OAVELKVASALDADVYKNEVDANVDEHFMVETDRLLAQVYVNDGFTGPTSGMVS
VYTSILTKVAEVTQATVDRYVETGLEASVSPERKRVVYVSDVPTDAILMTALORL
TDVSAQSFDDLAAHENGMDRMOQSDIOIGEDTEAQGRPNLFLPSTYYGDDPR"

```

Alignment Scores:

Pred. No.:	Length:	Matches:	Mismatches:	Indels:	Gaps:
Score: 217.50	302050	103	53	156	16
Percent Similarity: 39.90%					
Beet Local Similarity: 26.34%					
Query Match: 9.75%					

US-10-069-062-7 (1-432) x AL935257 (1-302050)

Qy 7 AAlaProGlyValAlaPheLeuAlaGlyValLeuValLeuGluProIleTyrAspAla 26
 DB 17016 GCCCTGGAACCTCAATTCGCCGTGAATACGCCGTAGTGAAGAGTGTATTCAGCA 16957

QY 27 TyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLysGlyThrSerLeu 46
 Db 16956 ATATTGTTGCACTTAATCAATTTGTGACGGCAACATTTCACTT----- 16912
 QY 47 LysGluSerArgIleLysIleSerSerProGlnPheAlaAsnGlyIuTrpGlyTrHis 66
 Db 16911 AGCGAATACTTGGAGCATTCGTTTGTGACGATCAAGAAACTCAATCGTCTGGCA 16852
 QY 67 IleSerSerAsnThrGluLysProArgIuValGlnSerArgIleAsnProPheLeuGlu 86
 Db 16851 CGTCAGGAGACCGC-----ANGCTTTGATTAATGTCACATCCCTTC----- 16807
 QY 87 AlaThrIlePheIleValLeuAlaTrpIleGlnProThrGluAlaPheAsnLeuGluIle 106
 Db 16806 -----CATATATTTTACGACGATCAATCTGACGAAGATTATGCCATGATTA 16756
 QY 107 IleIleTyrSerAspProGlyTrHisSerGlnLysAspThrGluThrLysThrSerSer 126
 Db 16756 ----- 16756
 QY 127 AsnGlyLysThrPheLeuTrHisSerArgAlaIleThrGluValGlu----- 143
 Db 16755 ---GGCGTGAACCTTGCTTTATCACTCGGTATCAATAGCGAATCGATGCTGAT 16699
 QY 144 -----LysThrGlyLeuGlyLysSerAlaGlyLeuValSerValAlaThrSerLeu 161
 Db 16698 GGTAAATAATAGGTCTCGGCTCTTCAGACGA-----GTACCGTGGCAACCGTCAAGCA 16642
 QY 162 LeuSerHisPheIleProAsnValIleSerThrAsnLysAspIleLeuHisAsnValAla 181
 Db 16641 CATATGCACTTTTACCATTTACCGATGATTAATAA-----CTGTCAATATGACGC 16588
 QY 182 GlnIleAlaHisCysTyrAlaGlnLysLysIleGlySerGlyPheAspValAlaThrAla 201
 Db 16587 GCGATTGCCATCTATCCGTGCAAGGCAAT---GGTTCCTTGCTGATCACTCTGCCAC 16531
 QY 202 IleTyr---GlyLeuIleValTyrArgArgPheGlnProAlaLeuIleAsnAspValPhe 220
 Db 16530 GCTATGATGGTGGATGCGCTACCATTCATTCGATCGGCANGTGTACAC----- 16480
 QY 221 GlnValLeuGluSerAspProGluLysPheProThrGluLeuLysLysLeuIleGluSer 240
 Db 16479 -----CTACAACAAGCTCACTCAAGCTTAAGTATGTTGACAAATG 16438
 QY 241 AsnTrpGluGlu---LysIleGluArgCysThrLeuProTyrGlyIleLysLeuLeuMet 259
 Db 16437 CCGTGCCCGACCTAAATAATCGATTACTGACCCCTCCAGACCGCTTACGTTGATGAT 16378
 QY 260 GlyAspValLysGlyLysSerGluThrProLysLeuValSerArgValLeuGlnTrpLys 279
 Db 16377 GGTGGACGGGTTCACCGGCTCCACATCACTTAGTGCATTAAGTGGCTTACTGCAAG 16318
 QY 280 LysGluLysProGluLysSerSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGln 299
 Db 16317 GCGAAGCAACGACGCA-----TACCAAGACCTTTCTAAAGCCAGCAAGCT 16270
 QY 300 PheMetLysGluLeuArgIuMetArgIuLysTyrAspSerAspProGluThrTrpIle 319
 Db 16269 TGTCTAAACGTATG-----ATC 16252
 QY 320 LysGluLeuAspHisSer---ValGluProLeuThrValAlaIleLysAsnIleArgLys 338
 Db 16251 ACTGTTTTCATCAGATGATCTGCACTGATTTCAAACCGAGTTCGCAAGATGCTCA 16192
 QY 339 GlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspValGlnThrGln 358
 Db 16191 CTACTTCGAAATTTGGCGGTTTCACTGACGTCAGATGAA-----ACACCC 16144
 QY 359 LeuLeuAspArgCysGlnGluIleProGlyCysValGlyGlyVal-----ValProGly 376
 Db 16143 CTGCTGACGCGTATGATTAATCTCGCCGAAGTGTGCGGCGCGCAAGACTTGGGC 16084

QY 377 AlaGlyGlyTyrAspAlaIleAlaValLeuVal 387
 Db 16083 GCTGCTGGTGTGATCGTGGATCGTTTGATT 16051
 RESULT 23
 AF290093
 LOCUS
 DEFINITION
 Enterococcus faecalis mevalonate kinase (mwak1), mevalonate
 diphosphate decarboxylase (mwad), and phosphomevalonate kinase
 (mwak2) genes, complete cds.
 ACCESSION
 AF290093
 VERSION
 AF290093.1
 KEYWORDS
 SOURCE
 ORGANISM
 Enterococcus faecalis
 Enterococcus faecalis
 Bacteria; Firmicutes; Lactobacillales; Enterococcaceae;
 Enterococcus.
 REFERENCE
 1 (bases 1 to 3152)
 Wilding, E.I., Brown, J.R., Bryant, A.P., Chaliker, A.F., Holmes, D.J.,
 Ingraham, K.A., Iordanescu, S., So, C.Y., Rosenberg, M., and Gwyn, M.N.
 Identification, evolution, and essentiality of the mevalonate
 pathway for isopentenyl diphosphate biosynthesis in gram-positive
 cocci.
 JOURNAL
 J. Bacteriol. 182 (15), 4319-4327 (2000)
 MEDLINE
 20353468
 PUBMED
 10894743
 REFERENCE
 2 (bases 1 to 3152)
 Wilding, E.I., Brown, J.R., van Horn, S., Mathie, T., Warren, R. and
 Gwyn, M.N.
 Direct Submission
 TITLES
 Submitted (26-JUL-2000) Department of Microbiology, SmithKline
 Beecham, 1250 S. Collegeville Road, Collegeville, PA 19426, USA
 FEATURES
 location/Qualifiers
 1..3152
 /organism="Enterococcus faecalis"
 /mol_type="genomic DNA"
 /db_xref="taxon:1351"
 104..1048
 /gene="mwak1"
 104..1048
 /gene="mwak1"
 /note="mwak1"
 /codon_start=1
 /transl_table=11
 /product="mevalonate kinase"
 /protein_id="AA02440.1"
 /db_xref="GI:9937386"
 1048..2043
 /gene="mwad"
 1048..2043
 /gene="mwad"
 /note="mwad"
 /codon_start=1
 /transl_table=11
 /product="mevalonate diphosphate decarboxylase"
 /protein_id="AA02441.1"
 /db_xref="GI:9937387"
 2046..3152
 /translaction="MISGKARAHYNALIKTWGRANBEYILPMNNSISLTIDATYET
 TVTDAHVSDEVFTLNGILONEKOTKVEKLELVROQADCTWPAKVESQNFVPAAG
 LASASGIALAAGACVNLGINTLSAKDLRLARGSGSACSIFGFGQNNKGGDET
 SFAENIPANWENELAMFLINDGERDVSRLDMKRTVETSSFYOGMLNVEKDLQ
 VHEAIKTKDPRLEITLHANGLRHGGTTIGAVPPTWSPGSLQAMLVROARAKGIP
 CYFTMDAGPNVTVLVEKKNLALTKTPISEHSKQLVPAPAGPQIELEFTGMCK"
 2046..3152
 /gene="mwak2"
 2046..3152
 /gene="mwak2"

```

/!note="WvAK2"
/codon_start=1
/translation_table=11
/product="phosphomethyltransferase kinase"
/db_xref="GI:9937388"
/protein_id="A602442.1"
/translation="MIETVTPGKLFAGEVAVVEPGRPAITVAVDQFTVTEETDE
GSIQAOYSLPIPRTRNGELVDIRNPFFHYLAHLEIKYAEQKELSFYILK
VTSELDSSNGKRYGSSGAVTGVKALNIPYDLINEMPEIKLSALAVQNGS
CGDIASCYGKWIAPSTPDHDMNCKVTETLIDLANMPEIMPLFVAPDRLI
GNTGSPASTSDLYDRVHOSKEEKAAVEGLMKSRICVETMNGEFTGKISYOKIT
KNRQLAEISLIVLETALEKLNLCIDLASVYGAASSGAGGDCGIVIPKSGIL
PLMTAEKKGIDPLPLHYVYTGKCKEKHESKR"
BASE COUNT      1008 a      591 c      670 g      883 t
ORIGIN
Alignment Scores:
Pred. No.:      3,66e-08      Length:      3152
Score:          209.50      Matches:      113
Percent Similarity: 39.38%      Conservative: 52
Best Local Similarity: 26.97%      Mismatches: 137
Query Match:    9.39%      Indels:      117
DB:              1          Gaps:      23
US-10-069-062-7 (1-432) x AF290093 (1-3152)
QY      6 SerAlaProGluYsaAlaPheLeuAlaGlyYrLeuValLeuGluProIleTyraP 25
      2058 ACTACGCCGAAAGTATTATTATGACGAAATAGCCGTGTGAACTGACGCCACCT 2117
      26 AlaTyValThraLeuSerSerArgMetHisAlaValIleThrProGlySer 45
      2118 GCCATTATGCTTCTGTGGAT-----CAATTGTAAGTACTGTGCAAGAACACA 2171
      46 LeuYseGluSerArgIleYsaIleSerSerProGluPheAla----- 59
      2172 GATGAAGGCACT-----ATTCAATCTGCAACATACAGCTTTACTTATCGTTG 2222
      60 -----AsnGlyGluTrpGluYrHisIleSerSerAnthrGluysProArgGlu 76
      2223 ACACGCCGAAATGTGAGCTCGATTAGATT----- 2255
      77 ValGlnSerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTy 96
      2256 -----CGCGAAATATCTTT-----CATATGTTCTAGCGGCGAT 2291
      97 GlnProThrGluAlaPheAspLeuGluIleIleIleTySerAspProGlyYrHis 116
      2292 CATCTTAAGTCAAAATAT-----GC 2312
      117 GlnGluAspThrGluThrIleThrSerSerAsnGlyGluYsrhrPheLeuYrHisSer 136
      2313 CAAGAG-----CAAAACAAGAAATGTCATT--TATCAATT 2348
      137 ArgAlaIleThrGluValGlu-----LysThrGlyLeuGlySerSerAla 151
      2349 AAGTAGCAGATGATAGTACTGCAATGACGAAATATGCTGCTGTTCAAC-- 2405
      152 GlyLeuValSerValAlaIleThrSerLeuSerHisPheIleProAsnValIleSer 171
      2406 GGTGACCTTAACCTTGAACTGCAAGCCTTGAATTTTATATGAC-----TAA 2459
      172 ThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysYrAlaGlnYs 191
      2460 TTGGAAATAGGAAATTTCAATATATCGCAATTAGCTACTTACCGCTTCAAGAAAT 2519
      192 IleGlySerGlyPheAspValAlaThrAlaIleTy-----GlyLeuIleValTy 210
      2520 ---GGTTCTTGGGAGATATCGCGCAGCTGTTAGCGGCGCTGAGATTGCTTTCAAC 2576
      211 PheGlnProAlaIleuIleAsnAspValPheGlnValLeuGluSerAspProGlyYr 230
      2577 TTGATCATCATGATTGGTCAAT-----CAAAAAGTA 2606

```

```

QY      231 ProThrGlu---LeuYseYsaIleGluSerAsnTrpGluGlu----- 244
      2607 ACCACTGAAACATTAAGTATTGTTAGCAATGACAGCTGCCGAAATATATGATTTTCCG 2666
      245 ---LysHISgluArgCysThrLeuProTyGlyIleYsaLeuLeuMetGlyAspValYs 263
      2667 TTTAAA-----GTACCGAACAACATCGTTTACTAATGGTTGACAGCT 2711
      264 GlyGlySerGluThrProYsaLeuValSerArgValLeuGlnTrpYsaYseGluYsPro 283
      2712 AGTCTCGCTGCACCTTCAAGACTAGTATGATGATGATGATCAATCAAAAGAAAGAAA--- 2768
      284 GluGluSerSerValValTyraPglInLeuAsnSerAlaAsnLeuGlnPheMetYsGlu 303
      2769 -----CAAGGCGCTTATGACAGCTTTTATGAAA 2798
      304 LeuArgGluMetArgGluYsaPseAspProGluThrTyrlleYseGluLeuAsp 323
      2799 AGTCGGCTTTGT-----GTGCAAAACAATGATTAT-----GCG 2831
      324 HisSerValGluProLeuThrValAlaIleYsaAsnIleArgYseGlyLeuGlnAlaLeu 343
      2832 TTTAACACAGAAATATTCTGTATTCAAAACAATTAATAATCCCAATTGCTC 2891
      344 ThrGlnYseGlu-----ValProIleGluProAspValGlnThrGlnLeuLeu 360
      2892 GCCGATATATCTTCACTAGCTGTGTGTAATCGAAACGAAGCCTGAAAAATCTTCT 2951
      361 AspArgCysGlnGluIleProGlyCysValGlyValValProGlyAlaGlyYr 380
      2952 GATTGGCTGATGATCTTATACAGAGCTGGAATCT-----TCGGCGCTGCGGCGGC 3005
      381 AspAlaIleAlaValLeuValLeuGluGluGlnValGlyAsnPheYseGlnYsrhr 399
      3006 GATTGTGGATGATTAT-----TCCGCCAAAATCT 3038
RESULT 24
AE016949/c 300225 bp DNA linear BCT 28-MAR-2003
LOCUS
DEFINITION
Enterococcus faecalis V583, section 3 of 11 of the complete genome.
VERSION
AE016949.1 GI:29342733
KEYWORDS
SOURCE
Enterococcus faecalis V583
ORGANISM
Enterococcus faecalis; Lactobacillales; Enterococcaceae;
Enterococcus.
REFERENCE
1 (bases 1 to 300225)
Paulsen, I., Banerjee, L., Myers, G.S.A., Nelson, K.E., Seshadri, R.,
Read, T.D., Fouts, D.E., Eisen, J.A., Gill, S.R., Heidelberg, J.F.,
Tettelin, H., Dodson, R.J., Umayam, L., Brinkac, L., Beanan, M.,
Daugherty, S., Deboy, R.T., Durkin, S., Kolonay, J., Madupu, R.,
Nelson, M., Vamathevan, J., Tran, B., Upton, J., Hansen, T., Shetty, J.,
Khouri, H., Utecherack, T., Radune, D., Ketchum, K.A., Dougherty, B.A.
and Fraser, C.M.
Role of Mobile DNA in the Evolution of Vancomycin-Resistant
Enterococcus faecalis
Science 299 (5615), 2071-2074 (2003)
JOURNAL
PUBMED
12663927
2 (bases 1 to 300225)
Paulsen, I., Banerjee, L., Myers, G.S.A., Nelson, K.E., Seshadri, R.,
Read, T.D., Fouts, D.E., Eisen, J.A., Gill, S.R., Heidelberg, J.F.,
Tettelin, H., Dodson, R.J., Umayam, L., Brinkac, L., Beanan, M.,
Daugherty, S., Deboy, R.T., Durkin, S., Kolonay, J., Madupu, R.,
Nelson, M., Vamathevan, J., Tran, B., Upton, J., Hansen, T., Shetty, J.,
Khouri, H., Utecherack, T., Radune, D., Ketchum, K.A., Dougherty, B.A.
and Fraser, C.M.
Direct Submission
Submitted (03-FEB-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..300225

```

gene /organism="Enterococcus faecalis V583"
/mol_type="genomic DNA"
/strain="V583"
/db_xref="taxon:226185"
102..557
/gene="argR"
/locus_tag="EF0676"
102..557
/gene="argR"
/locus_tag="EF0676"
/note="similar to GP:4127534, GB:M28829, and PID:551976;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="arginine repressor"
/protein_id="AA080498.1"
/db_xref="GI:29342734"
/translation="MKAKRQLIKQLIQKEIFETODELITLBEIGVATQATVSRD
IREMSIVKTHGADGVKAIIFSQAQSTSEELRSKDVSMRMRVQPIVILHTEMG
NADVSNFIDEVAYPEVAGTVAGADTIIITRSEBDHFIERIENMTF"
591..2099
/locus_tag="EF0677"
591..2099
/locus_tag="EF0677"
/note="similar to GP:9755778; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphoglucomutase/phosphomannomutase family
protein"
/protein_id="AA080499.1"
/db_xref="GI:29342735"
/translation="MAELKALQNGSDIRGIALDTREQNTLPTATAVAEIVGVNMLQ
DKQIPRAQOQLTIAIGHDSRLTAEPKQALVDTPLSIGIOTVDVGLATPAMMAT
OPTIQCDAIATITASHLPYRNGIKFTABEGAKEDIRITLSTDELTANENCTLM
KQELPITAEHLVEKIRQGISPEEKPIQGFRIIVDAGNGAGGFAEAVQLVLAGDVT
GQFLPDGHPNHNVPDENSEAMKSIQAVLANQADIGIIPDIDRVSAAVDDGCV
LARNMLAVTAAIVLEKAPSYITNSPTSHKTPIEBKGGQAIRYSAGYVINKM
IEINHGFGOTPLAETSGHAFQENYNDLDGAVVAKTLMILPELKNONOTIGDLIAT
LKQAPETREFRKTAEDVTCYGCQVLDPELFVFNQDPFANDRENQSGVRNVSQY
GSGWFLRLSLHPELIVLOVENDQSKNACVLEKATPLQKTEEIDSQIEK"
2123..2677
/locus_tag="EF0678"
2123..2677
/locus_tag="EF0678"
/note="similar to GP:10174740; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="acetyltransferase, GNAT family"
/protein_id="AA080500.1"
/db_xref="GI:29342736"
/translation="WENVIKGKGVGFSHREYIYTHIAKQOMDNELHLHSLMDALHPM
GLEWKPDTINKGDDRLFLALENTTEPIGWVLSVQLKNRGNANIGAILQKEOR
GGCGFEAVSLICKPAFELGLHKIRLAVNSNNQAIHYEKVGFKEGIDREALFOD
GQMLDIYNGIIGKEWLMQIKAES"
complement(2803..3666)
/locus_tag="EF0679"
/note="This region contains a pseudogene, one or more
premature stops, and is not the result of a sequencing
artifact; ribosomal large subunit pseudouridine synthase,
ruid subfamily; similar to GB:X74862, and PID:510996;
identified by sequence similarity; putative"
4009..6195
/locus_tag="EF0680"
4009..6195
/locus_tag="EF0680"
/note="similar to GP:292644; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="penicillin-binding protein 2a"
/protein_id="AA080501.1"

gene /db_xref="GI:29342737"
/translation="MDNLKQFSSKVGGLRHFMTWIKRYLIQFHQARRKIWKQYINK
ILILGLVVALGASIVLYFLAKSANVETLKSLSSTVYDSGKVGGLFQKGFV
ELDNISPIYQDAVISTEDRGFQHKGSIKGIARAVVGLTKFGKIGGGGSGTITQOL
AKNAVLTQEDLDRKARELPLAIEIKKYSKKDILAMTLANSYFENGWGVQDAARX
FGVDASQVTEBAATLAKMLKPGIYNIDVINDATARNVTLQMLVNRKLSQSEAN
OASVNLASLNDYVGDENSKYPPYDAVIDAVNKKPEEDILNKGYKITSLN
OGYODAMATYKNDALFPNADGMAVQSGAIDPKGQVALVGRGEHYRGFNL
ATQTKRSPGSLKPISTVYTPALBAGKVSVDLPDQDYKAPONSRTYSGEVMYOA
LGEISLPAVWLHLDIGDKYEKTEKNGIPLSEKDRYYGLALGSLQTCVSPSTWAGA
YSAPANEKYTEHLITKIVSTGAIVDNTRKVEQVITKQVADGITSMLGVSSG
GSAVAPAGVYMAKVTGTETNFDSSKTNDDWVIGTEBVVATMLGQETSHTYLE
GSSATYASQVNSQASGILPVQKQFPAVAAVYAGCVNDSNLTGANSNMRDNL
KDFNSVDDGASNFDDTINVEGQIKDKAKKVKDKIEGIFGLLGN"
6258..6599
/locus_tag="EF0681"
6258..6599
/locus_tag="EF0681"
/note="similar to GP:3688819, and GP:3688819; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA080502.1"
/db_xref="GI:29342738"
/translation="MSNIYPTANOIERELREDEPKALQAAVEYKADEAAVTLPEK
QGFQGLQEKMRGEEFTDEDAEKQSIATVQETAVINELMKQKQASVINDLRI
IMTPRYMDND"
6755..7990
/locus_tag="EF0682"
6755..7990
/locus_tag="EF0682"
/note="similar to SP:P80102; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA repair exonuclease family protein"
/protein_id="AA080503.1"
/db_xref="GI:29342739"
/translation="WKLHTADLMDNSFESLTGIPRLARLRANQQLMNTVYTA
IREVNDLVPAGDTPHQSQTSISMOQALMALEQKQKEIVITLFGHNDYKKDORV
FSPDNVFLPEKWEVETLYPETKAGQVAVSGFYEHPHIDENALPEPIYQASDII
IGIYHGDTSRAQOQNVAFPTWKDKATGYNWMALGHIPQIVSEQIIVYGPFGH
TKKQSLQGVAVVTLSQNAVQPEKAEIDTMEBVSALACPDQSVLSYETSILT
KMOAHQOQOQMALTLKXCHQLDVTVQAVUNGSELLSYOQQLQKQDPPVYVLILO
AEEPTKEPIYLAASPMLITALEKTYLDVPIEDTTKELRIPEFAGITSMDQMRLES
LRLAEQINEDFTIQEDPL"
7987..10698
/locus_tag="EF0683"
7987..10698
/locus_tag="EF0683"
/note="similar to SP:P80102; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA080504.1"
/db_xref="GI:29342740"
/translation="WKLAVAVIVGFGKQOQKITFESGNQLLYGANEVKSITLYOFIQ
AMLFGPPLKGRKADYAPKVGSGYGLMLHPHAFGEVAFPEKNGQAIYYQOQ
MDEKTKIQOMLHLPTKLFQBFPOQBOULITLBEELQSLAILVLAGFLPSLP
YRNAYPEAKOITFGKGOQPLNKLQAVOELKXOIEKQOQOFPQOLETRETQO
ORVTLQOIKRETQODLQVABQORHPFLYEWOSLPRETNVTVKQNOLETFEEO
YMYDKERQRLTKELSLQSEALDVPBGATPLBQBALIQLLNRYDMQQLMTESEM
TQTEQRQREVNLTLENEMSWSPERPOLPVDNETVQKROQVAVYQALQGGQSKAF
LASQIENLEQDLPFERKHAAPFGEATRAQAPQSOPLMILASITLVLAGFLPSLP
KMGVLVGVSAIVAKVSYATRKASDVEYEQTKLSQDLYNBEQOQALNDBREA
OLOVEQLEAVABQAVOAGQOQIDTLMNORRELITRYLLVOTNEELKQLEKNOQ
KROVREBOIRPLMTHLPLQKSKSEKHLFEGFEMEKVPAFQYQADVTLKOQRE
LKTQOQALTRIPULIRILSLIADVPSTRIQALITQOQOVQSGQELATQUGLPPA
VTEVTLDRSLQALQOQALEBSQDKIQOQOVALTYKQQLMTGDTLDELVQRAILK
AEIKELQWSGQOLAGOLIMDLITELSEBOVFLVAYSYFPLMLNOROSIOVAB
GOLVAVTKQEMFVHELSTGTQQLMMNAVFALVQGEQLICPIIIDDGMHVDHQ

QY 392 Val 392
 Db 3004 ATT 3006
 RESULT 26
 LOCUS LHE279018
 DEFINITION Lactobacillus helveticus 53/7 gene cluster for the Mevalonate
 Pathway including mvk gene, mdp gene, pmk gene and cbp gene.
 ACCESSION AJ279018
 VERSION AJ279018.1 GI:15212069
 KEYWORDS carotenoid biosynthesis protein; cbp gene; mdp gene; mevalonate
 diphosphate decarboxylase; mevalonate kinase; mvk gene;
 phosphomevalonate kinase; pmk gene.
 SOURCE Lactobacillus helveticus
 ORGANISM Lactobacillus helveticus
 Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
 Lactobacillus.
 REFERENCE 1
 AUTHORS Smeds A., Purtsel T. and Palva A.
 TITLE Identification of a gene cluster for the mevalonate pathway in
 Lactobacillus helveticus
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 4350)
 AUTHORS Smeds A.
 TITLE Direct Submission
 JOURNAL Submitted (11-SEP-2000) Smeds A., Basic Veterinary Sciences,
 Faculty of Veterinary Medicine, P. O. Box 57, 00014 University of
 Helsinki, FINLAND
 FEATURES
 source
 1. 4350
 /organism="Lactobacillus helveticus"
 /mol_type="genomic DNA"
 /strain="53/7"
 /db_xref="taxon:1587"
 129..134
 -35 signal
 -35 signal
 -10 signal
 -10 signal
 156..161
 190..195
 204..209
 218..1126
 /gene="mvk"
 /gene="mvk"
 218..1126
 /function="essential for the mevalonate pathway"
 /codon_start=1
 /transl_table=11
 /product="mevalonate kinase"
 /protein_id="CAC51370.1"
 /db_xref="GI:15212071"
 /translation="MRTSRPLAHGKVIILIGHSVYVGDALAMPKALHKTIVYENPE
 MMDDTARHGSPFDTPSEIDGLKYVATKQKANNBPLKITITGEIPMERGSSAA
 VALGTTKALNQPSLNMSESEIMAVTNHAKINHGKASGDAATVNSDVLVFNKKMG
 PKVLQAGLGAITLPIMDTIGELGTRKKAQAQVSLSEBQKAMKSLGLADLTCKXAG
 INHDVTVGNI FNEAQKILHSFNISTTKIDQLQKIALDNGALGFKLSGGGLGSTITAL
 CESQGLAENIAENACODLISNYWYBEI"
 1116..1121
 /gene="mvk"
 /gene="mvk"
 1126..2088
 /gene="mdp"
 1126..2088
 /gene="mdp"
 /codon_start=1
 /transl_table=11
 /product="mevalonate diphosphate decarboxylase"
 /protein_id="CAC51371.1"
 /db_xref="GI:15212071"
 /translation="MSKIVRAHNTNIALIKYWKANAEIRLPLMSSLSMTLDAFYDTS
 VKTRBNATFYINGCKQDSKQSORVPSYTLKQNKRGTHDNLIVKSNHYPTPSAS
 SSSAFALAAAFQCYINI QVKEKLSRLAIGSACRSIFGFSWQKQDSASSYA
 VALDEHPQDLIAVELINTQKISTSGMKRQSPFPNPLMBENETLNMQIMAI
 KNDPFALGELAEINAMEMHAINLTAQPKTFYEPPTIRAIKLVEDLRKEGIECYTI

RBS
 gene
 CDS
 DAPNIIKLCRLNSKEITERFKSVNNVNIIVANFGPITCLD"
 2319..2322
 2330..3214
 /gene="pmk"
 /gene="pmk"
 2330..3214
 /gene="pmk"
 /codon_start=1
 /transl_table=11
 /product="phosphomevalonate kinase"
 /protein_id="CAC51372.1"
 /db_xref="GI:15212072"
 /translation="WVIDRNDPFEYILSAISYEOYCIEONIKNKQVYDLHANSDLDS
 ADGKXYGLSSAAVAVATVKAILHFGYVMSNELVYKLSAISHSYVQSGAGDIAAS
 VYGMVAYOTFPKMWLOYELAKTISDVYVNEPMPGKILPLPMDKLMIGSGOKPA
 SPRLVDENRANKALNTEYKRFALASSSCVAKMVGPAKNIALIKQIORYNRHILQ
 HPAKINQALIEIPRLITQIKIADPFGAAKTSGAAGNDCGIYITDADVDVLAENMR
 RNLGILPNRVRHQLSLAH"
 3217..3222
 3227..4243
 /gene="cbp"
 3227..4243
 /gene="cbp"
 /codon_start=1
 /transl_table=11
 /product="putative carotenoid biosynthesis protein"
 /protein_id="CAC51373.1"
 /db_xref="GI:15212073"
 /translation="MSISERKEEHLKLAQMPFKKNSPDMHLLPRLPEKVDQ
 SVLATEMFNKSADPFIAMTGGSKQSLIVNQLAKLHOKRIALNLSASITLAKK
 DLESFPYARDSDANGVLI VNNPPTPIALIKQIKELQADLQHLNLTQASITLAKK
 DNNFIWLDQIKNILDQITIPVLIKESVGLDONSJHLKENGIKFPDVAGSGGIFAQ
 ENRANRHDVSYLEDIGLPTVISALMAQESVNFVSGVRPLDILKGLSGQFPVG
 IENVPLQERNKNSFEGLOTLLISNMWDELAALAIYIGKIDLASLPKIKKYVDLPKNQI
 DDLTI"
 3252..3283
 /gene="cbp"
 4248..4280
 terminator
 BASF COUNT 1455 a 739 c 798 g 1358 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,63e-06 Length: 4350
 Score: 190.00 Matches: 103
 Percent Similarity: 38.86% Conservative: 75
 Best Local Similarity: 22.49% Mismatches: 164
 Query Match: 8.52% Indels: 116
 DB: 1 Gaps: 24
 US-10-069-062-7 (1-432) x LHE279018 (1-4350)
 QY 7 AIAProGLyVAlAphELeuNlaGlyGlyTyrLeuValLeuGluProTleTyrAaPa 26
 Db 2147 GCACCCGGAATTTATACATCGCTGGAGAAACGCCGTTTGGACAAATTTCTCTGCT 2206
 QY 27 TyrValThrAlaLeuSerSerArgwethEsaIaValIleThrProLySgLyThrSerLeu 46
 Db 2207 GTTTAGTTCGCCGTAACCAATTGACGTGTTCAATTCAATAAAGCAAGCTTCTACT 2266
 QY 47 LysGluSerArgIleLysIleSerSerProGlnPheAlaSerGly-----GluTrpGlu 64
 Db 2267 GGT-----TTGATTCACTCCAAACAAATATTCACAAAGATTCTATTGATGGGTC 2314
 QY 65 TyrHisIleSerSerArgThrGluLysProArgGluValGlnSerArgIleAsnProPhe 84
 Db 2315 CGTCAGGCTCTAAATGTA-----ATTGATACCGTGATTAACCATTT 2359
 QY 85 LeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnProThrGluAlaPhe----- 102
 Db 2360 -----GAATATATTATCTGCAATTTCTTATACAGAAACAATACGCAAT 2404
 QY 103 -----AspLeuGluIleIleIleTyrSerAspProGlyTyrHisSerGlnGluAspThr 120
 Db 2405 GAACAAATATCAAGATGAAGTATATGAT-----CTGCATGTTAATCAGATCTA 2455

Db 108362 -----CTGAAGTTCACCAAACTGACCTGGGTGCATCAGCT 108327
 QY 152 GlyLeuValSerValValAlaThrSerLeuLeuSerHisPheLeu----- 166
 Db 108326 GCCATGACCAACATCAGTATGTGACGCTCTTCTT---CATTATCTTGCAACTGTTAACTT 108270
 QY 167 -----ProAsnValIleSerThrAsnLysAspIleLeuHis 178
 Db 108269 TCATGTTTGGGCCAATCTCCGGTGGCAACGACGCTGACAGATCTTGACTTCTGTCAT 108210
 QY 179 AsnValAlaGlnIleAlaHisCysTyrAlaGlnLysLysIleGlySerGlyPheAspVal 198
 Db 108209 GCTATTGCTCAAGATGACATGTGTATGACACAGAAATAATGGCAGTGGGTTGATGTT 108150
 QY 199 AlaThrAlaIleTyrGlyLeuIleValTyrArgArgPheGlnProAlaLeuIleAsn 217
 Db 108149 AGTGTGCTGTCTATGGAGTCAACGTTACACAAGGTTTCTCCAGAAATATATATCC 108093

RESULT 28

LOCUS AX142195 1077 bp DNA linear PAT 31-MAY-2001
 DEFINITION Sequence 917 from Patent WO0134809.
 ACCESSION AX142195
 VERSION AX142195.1 GI:14281969

KEYWORDS
 SOURCE synthetic construct
 ORGANISM artificial sequences.

REFERENCE
 1 Kimmerly, W. J.
 Staphylococcus epidermidis nucleic acids and proteins
 Patent: WO 0134809-A 917 17-MAY-2001;
 JOURNAL GLAXO GROUP LIMITED (GB)

FEATURES

Source 1..1077
 1. /organism="synthetic construct"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"
 /note="synthetic nucleic acid sequence"
 Location/Qualifiers
 BASE COUNT 368 a 158 c 219 g 332 t
 ORIGIN

Alignment Scores:

Pred. No.: 2.27e-06 Length: 1077
 Score: 181.50 Matches: 99
 Percent Similarity: 36.74% Conservative: 52
 Best Local Similarity: 24.09% Mismatches: 131
 Query Match: 8.14% Indels: 129
 DB: 6 Gaps: 18

US-10-069-062-7 (1-432) x AX142195 (1-1077)

QY 7 AlaProGlyLysAlaPheLeuAlaGlyGlyTyrLeuValLeuGlnProIleTyrAspAla 26
 Db 16 GCCCCCGGAAACCTTATATATGACGGCAGATATGACGATACGAAACGAGATATTAATCT 75
 QY 27 TyrValThrAlaLeuSerSerArgMetHisAlaValIle----- 39
 Db 76 ATTCTATATGCAAGTAAATCGCTTGTAAACGCGACCAATGAGCGCTCAATAAAGTTGAA 135
 QY 40 -----ThrProLysGlyThrSerLeuLysGlnSer 49
 Db 136 GGTAGTATTCATTCACAAACATTAACATTATGAAACCACTTAATTAATGACCTATGAAAGT 195
 QY 50 ArgIleLysIleSerSerProGlnPheAlaAsnGlyGlnTyrGlnTyrHisIleSerSer 69
 Db 196 AGAATTCGAATCTCAGATGTTCAAGTGCCTAAGCAA----- 231
 QY 70 AsnThrGlnLysProArgGlnValGlnSerArgIleAsnProPheLeuGlnAlaThrIle 89
 Db 232 -----CTGAATATATGTTGTG 246

QY 90 PheIleValLeuAlaTyrIleGlnProThrGlnAlaPheAspLeuGlnIleIleTyr 109
 Db 247 ACAGCATATGAGAGTGTGACATATGTCGACAGTGCATATGATTAATTA----- 297
 QY 110 SerAspProGlyTyrHisSerGlnLeuPheThrGlnLysThrSerSerAsnGlyGln 129
 Db 298 -----MAGCCTTTCATTTAAACATTTGATTAACCTTAACAGATTAACCTCTGCT 345
 QY 130 LysThrPheLeuTyrHisSerArgAlaIleThrGlnValGlnLysThrGlyLeuGlySer 149
 Db 346 -----CAGAAGTACGATTAAGCTTCA 366
 QY 150 SerAlaGly----LeuValSerValAlaThrSerLeuLeuSerHisPheIleProAsn 168
 Db 367 AGCGCGCGCTGTTTATGATCTGTTGT-----AAAGCTTGAAGAATATATATGCT 417
 QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
 Db 418 -----TTGGAATTTATCAAACTTATATTTAATAATTAGCTGTATTCACAAATATGAATTA 474
 QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyr---GlyLeuIleVal 207
 Db 475 CAAGTTTAAGTTTCATGTGGC---GATATGGCGTTAGTGTCTACAGTGGTGGCTTGA 531
 QY 208 TyrArgArgPheGlnProAlaLeuIleAsnAsp---ValPheGlnValLeuGlnSerAsp 226
 Db 532 TATAGTACGTTTGAC-----CATGACTGGGTGAACACAGCAATATGAAAGA--- 576
 QY 227 ProGlnLysPheProThrGlnLeuLysLysIleGlnSerAsnTyrGlnLysHis 246
 Db 577 -----ACATCGGTGAATGATGATGTTTGGAAAAAATTTGGCAGGCTTACAT 621
 QY 247 ---GluArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlyGly 265
 Db 622 ATCGAACCTTTACAAAGCTCCGAAAAATATGACAGCTTATGATGATGACGTGGTCCA 681
 QY 266 SerGlnThrProLysLeuValSerArgValLeuGlnTyrLysGlnLysProGlnGln 285
 Db 682 GCTTCTTCCACACTTACGTAGTGAAGGTCAACGTTTAA--- 723
 QY 286 SerSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGlnLeuArg 305
 Db 723 ----- 723
 QY 306 GluMetArgGlnLysTyrAspSerAspProGlnTyrTyrIleLysGlnLeuAspHisSer 325
 Db 724 -----TCAAGTCCAAAGTTTATGATGATTTTATGATCAATCT 762
 QY 326 -----ValGlnProLeuThrValAlaIleLys-----AsnIle----- 336
 Db 763 CATGCTGTGTGAAGAAGTTTATCCAGCTTTTAAACATTAATATATCAAAAGTGTGCA 822
 QY 337 -----ArgLysGlyLeuGlnAlaLeuThrGlnLysSerGlnValPro 350
 Db 823 AAGATGATACGTATTAACAGACGATATTTATCATCTATGATGATTAACGAACATCAGTTGA 882
 QY 351 IleGlnProAspValGlnThrGlnLeuLeuAspArgCysGlnGlnIleProGlyCysVal 370
 Db 883 ATTTGAACAGATTAAGCTTAATAAATTTATGTGATGTCGGTGAAGACAGCGTGGC----- 936
 QY 371 GlyGlyValValProGlyValAlaGlyGlyTyrAsp 381
 Db 937 GCTTCTAATAAATTCAGGTCTGTGTGTGCGCAT 969

RESULT 29

LOCUS AF290091 3169 bp DNA linear BCT 29-AUG-2000
 DEFINITION Staphylococcus epidermidis mevalonate kinase (mvak2), mevalonate
 diphosphate decarboxylase (mvad), and phosphomevalonate kinase
 (mvak2) genes, complete cds.
 ACCESSION AF290091
 VERSION AF290091.1 GI:9937378
 KEYWORDS

QY 286 SerSerValValTyrAspGlnLeuAenSerAlaAsnLeuGlnPheMetLysGluLeuArg 305
 Db 2725 ----- 2725
 QY 306 GluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHisSer 335
 Db 2726 -----TAGATCCAAAGTTTATGTCGATTTTATGATTCATCT 2764
 QY 326 -----ValGluProLeuThrValAlaIleLys-----AsnIle----- 336
 Db 2765 CAGCTGTGTAGAAAGTTTAACTCAAGCTTTTAAACTAATATCAAAAGGTGTTCA 2824
 QY 337 -----ArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValPro 350
 Db 2825 AAGATGATACGTATTAACAGACGATATTCATCTATGATACGATACGACGATCAGTTGAA 2884
 QY 351 IleGluProAspValGlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysVal 370
 Db 2885 ATGAAACAGATTAAGCTTAAATAATATGATGTCGTGTAAGAACACGCGTGC----- 2938
 QY 371 GlyGlyValValProGlyAlaGlyLysTyrAsp 381
 Db 2939 GCTTCTAAACTTCAGGTCTGTGTGTGCGCAT 2971

RESULT 30
 AF270256/c 4010 bp DNA linear BCT 01-AUG-2000
 LOCUS Staphylococcus epidermidis strain SRI clone step.1057a10 genomic
 DEFINITION sequence.
 ACCESSION AF270256
 VERSION AF270256.1 GI:9624167
 KEYWORDS

SOURCE Staphylococcus epidermidis
 ORGANISM Staphylococcus epidermidis
 Bacteria; Firmicutes; Bacillales; Staphylococcus.
 REFERENCE 1 (bases 1 to 4010)
 Kimmerly, W.J., Taylor, J. David, Nelson, A.J., Godlevski, M.M.,
 Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I.,
 Listenebee, S., Ashanti, C., Altschuller, G., Mam, L., Shepherd, N.S.,
 Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
 Furdon, P.J.

TITLE Transposon-mediated sequencing of the Staphylococcus epidermidis
 genome
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 4010)
 AUTHORS Taylor, J. David, Kimmerly, W.J., Nelson, A.J., Godlevski, M.M.,
 Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I.,
 Listenebee, S., Ashanti, C., Altschuller, G., Mam, L., Shepherd, N.S.,
 Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
 Furdon, P.J.

TITLE Direct Submission
 JOURNAL Submitted (22-MAY-2000) Departments of Genomic Sciences and
 Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
 Drive, Research Triangle Park, North Carolina 27709-3398, USA

FEATURES
 source location/Qualifiers

1. 4010
 /organism="Staphylococcus epidermidis"
 /mol_type="genomic DNA"
 /strain="SRI"
 /db_xref="taxon:1282"
 /clone="step.1057a10"
 BASE COUNT 1230 a 806 c 556 g 1418 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.27e-05 Length: 4010
 Score: 181.50 Matches: 99
 Percent Similarity: 36.74% Conservative: 52
 Best Local Similarity: 24.09% Mismatches: 131
 Query Match: 8.14% Indels: 129
 DB: 1 Gaps: 18

US-10-069-062-7 (1-432) x AF270256 (1-4010)
 QY 7 AlaProGlyLysAlaPheLeuAlaGlyGlyTyrLeuValLeuGluProIleTyrAspAla 26
 Db 1156 GCCCCCGGAAACCTTAATATTCAGGCGAGTACGATACCGAACCGAGATTAATATCT 1097
 QY 27 TyrValThrAlaLeuSerSerArgMetHisAlaValIle----- 39
 Db 1096 ATTCTTATTCAGTAAATCGCTTGTAAACGGACGACATTTAGAGCGGTCAATTAAGTTGAA 1037
 QY 40 -----ThrProLysGlyThrSerLeuLysGluSer 49
 Db 1036 GGTAGTATTCATTCACAAACATTATACGATACCAAGTTAAATTCACCGTAAATGAAT 977
 QY 50 ArgIleLysIleSerSerProGlnPheAlaAsnGlyLysGluTyrGluTyrHisIleSerSer 69
 Db 976 AGAATGAAATCTCAGATCTTCAGCTGCTTAAGCAA----- 941
 QY 70 AsnThrGluLysProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThrIle 89
 Db 940 -----CTGAATATGTTGTC 926
 QY 90 PheIleValLeuAlaTyrIleGlnProThrGluAlaPheAspLeuGluIleIleTyr 109
 Db 925 ACAGCTATGAAAGTGTGTGACATATGCGCAGCTGACATATGATTA----- 875
 QY 110 SerAspProGlyTyrHisSerGlnGluAspThrGlnThrLysSerSerAsnGlyGlu 129
 Db 874 -----AACGCTTTCATTTAACTGATGATGATTAAGTACGATCTCTGCT----- 827
 QY 130 LysThrPheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLysGlySer 149
 Db 826 -----CAGAATGAGATTAGTTCA 806
 QY 150 SerAlaGlyLysValSerValValAlaThrSerLeuLeuSerHisPheIleProAsn 168
 Db 805 AGGCGCGCTGTTTATGATCTCTGTT-----AAAGCTTTGAACAAATGATGATGCT-- 755
 QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
 Db 754 -----TTGGAATTAACAACTTATATTTAAATTAATGCTGTAATGCAATATGAATTA 698
 QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyr-----GlyLeuIleVal 207
 Db 697 CAAAGTTTAAGTTCAATGTCG-----GATATGCGGATGATGTCATACGCTGTCGCTTGA 641
 QY 208 TyrArgArgPheGlnProAlaLeuIleAsnAsp-----ValPheGlnValLeuGluSerAsp 226
 Db 640 TATAGTACCTTTGAC-----CATGACTGGGTGAAACAGCAATATGAAAGAA-- 596
 QY 227 ProGluLysPheProThrGluLeuLysLysLeuIleGluSerAsnTrpGluLysHis 246
 Db 595 -----ACATCGGTGAATGATGTTTGGAAAAAATTTGGCCAGGCTTAACT 551
 QY 247 -----GluArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlyGly 265
 Db 550 ATCGAACCTTTACAGCTCCGAAAAATATGAAAGTCTTATGATGATGACGCTGCTCCA 491
 QY 266 SerGluThrProLysLeuValSerArgValLeuGlnTrpLysGluLysProGluGlu 285
 Db 490 GCTTCTTCCACACTTATGATGATGAGTCAAAACGTTTAAAA----- 449
 QY 286 SerSerValValTyrAspGlnLeuAenSerAlaAsnLeuGlnPheMetLysGluLeuArg 305
 Db 449 ----- 449
 QY 306 GluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHisSer 325
 Db 448 -----TCAGATCCAAAGTTTATGCGATTTTATGATTCATCT 410
 QY 326 -----ValGluProLeuThrValAlaIleLys-----AsnIle----- 336
 Db 409 CAGCTTGTGTAGAAAGTTTAACTCAAGCTTTTAAACTAATATCAAAAGGTGTTCA 350

QY 337 -----ArglysglyLeuGlnAlaLeuThrGlnLysSerGluValPro 350
 Db 349 AAGATGATACGTATAAAGACAGCATATTCATCTATGATACGAAAGATCAGTTGAA 290
 QY 351 ILeGluProAspValGlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysVal 370
 Db 289 ATGGAACAGATAGCTTAATAAATAATATGATGTCGCGTGAAGAACGACGCTGCGC----- 236
 QY 371 GlyIlyValValProGlyAlaGlyIlyTyraSp 381
 Db 235 GCTTCTAAACTTCAGGTGCTGGTGGTGGCAT 203
 RESULT 31
 AX145574 4010 bp DNA linear PAT 31-MAY-2001
 LOCUS AX145574.1 GI:14284139
 DEFINITION AX145574
 VERSION AX145574.1 GI:14284139
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct
 REFERENCE 1
 AUTHORS Kimmerly, W. J.
 TITLE Staphylococcus epidermidis nucleic acids and proteins
 JOURNAL Patent: WO 0134809-A 4296 17-MAY-2001;
 GLAXO GROUP LIMITED (GB)
 FEATURES
 source
 1..4010
 /organism="synthetic construct"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"
 /note="synthetic nucleic acid sequence"
 BASE COUNT 1230 a 806 c 556 g 1418 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,27e-05 Length: 4010
 Score: 181.50 Matches: 99
 Percent Similarity: 36.74% Conservative: 52
 Best Local Similarity: 24.09% Mismatches: 131
 Query Match: 8.14% Indels: 129
 DB: Gaps: 18
 US-10-069-062-7 (1-432) x AX145574 (1-4010)
 QY 7 AlaProGlyLysAlaPheLeuAlaGlyIlyTyraLeuValLeuGlnProIleTyraPala 26
 Db 1156 GCCCCCGGAAACTTATATTTGCGAGCGAGTATGACGTAACCGAACCGATATTAATCT 1097
 QY 27 TyrValThrAlaLeuSerSerArgMetHisAlaValIle----- 39
 Db 1096 ATCTTATGACGTAATATCGCTTTGTAACGCGACCAATGAGCGCTCAATAAAGTTGAA 1037
 QY 40 -----ThrProGlyIlyThrSerLeuysGluSer 49
 Db 1036 GGAATATTCATTCGAAACATTACATATGAAACCGATTAAATTGACCGTATGAAAGT 977
 QY 50 ArgIleLysIleSerSerProGlnPheAlaAsnGlyIlyTrpGluTyrisIleSerSer 69
 Db 976 AGAATGGAATCTCAGATGTTCAAGCTGCTAAGCAA----- 941
 QY 70 AsnThrGluLysProArgGluValGlnSerArgIleAsnProPheLeuGlnAlaThrIle 89
 Db 940 -----CTGAATATGTTGG 926
 QY 90 PheIleValLeuAlaTyrlleGlnProThrGlnAlaPheAspLeuGlnIleIleIleTy 109
 Db 925 AACGATATGAAAGTGTGTAACAGATATGCGCAGTTCGAAATATGATTTA----- 875
 QY 110 SerAspProGlyTyrisSerGlnGluAspThrGlnThrLysThrSerSerAsnGlyGlu 129

Db 874 -----AACCATTTTCATTTAAACCATGATAGTAACTAGACATACTCTGCT----- 827
 QY 130 LysThrPheLeuTyrisSerArgAlaIleThrGlnValGlnLysThrGlyLeuGlySer 149
 Db 826 -----CAGAAAGTACGATTAAGTTCA 806
 QY 150 SerAlaGly-----LeuValSerValValAlaThrSerLeuLeuSerHisPheIleProAsn 168
 Db 805 AGGCGCGCTGTTTATGATCTGTTGTT-----AAGCTTGATGAATATTCATGCT--- 755
 QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyraIa 188
 Db 754 ---TTGGAATATTAACCTTATATTTATTAATTATGATTAATGCAATATGAAATTA 698
 QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyra-----GlyLeuIleVal 207
 Db 697 CAAAGTTTAAGTTTCATGTGCG---GATATATGGGTTGATGCTACAGTGTGGCTTGCA 641
 QY 208 TyrArgArgPheGlnProAlaLeuIleAsnAsp---ValPheGlnValLeuGluSerAsp 226
 Db 640 TATAGTACGTTTGAC-----CATGCTGGGTGAACAGCAATATGAAAGAA--- 596
 QY 227 ProGluLysPheProThrGlnLeuLysLysLeuIleGluSerAsnTrpGluLysHis 246
 Db 595 -----ACATCGGTGAATGATGTTTGGAAAAAATTGGCCAGGCTTAACAT 551
 QY 247 ---GluArgCysThrLeuProTyrlleGlyIleLysLeuLeuMetGlyAspValLysGly 265
 Db 550 ATCGAACCTTTAACAAGCTCCGAAAAATATGAGAGTCCTTATGATGATGACGCGCTTCA 491
 QY 266 SerGluThrProLysLeuValSerArgValLeuGlnTrpLysLysGluLysProGluGlu 285
 Db 490 GCTTCTTCCACACTTACTAGTATGATGAGTCAACACCTTTAA-- 449
 QY 286 SerSerValValTyraSpGlnLeuAsnSerAlaAsnLeuGlnPheValGluLeuArg 305
 Db 449 ----- 449
 QY 306 GluMetArgGluLysTyraSpSerAspProGluThrTyrlleGlyLeuLeuAspHisSer 325
 Db 448 -----TCAGATCCAACTTTTATGCGATTTTATGATCAATCAATCT 410
 QY 326 -----ValGluProLeuThrValAlaIleLys-----AsnIle----- 336
 Db 409 CATGCTGTGTGAAAGTTTATCCAACTTTTAAACCTTAATCAATTAATCAAGAGTGTCA 350
 QY 337 -----ArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValPro 350
 Db 349 AAGATGATACGTATAAAGACAGCATATTCATCTATGATACGAAAGATCAGTTGAA 290
 QY 351 ILeGluProAspValGlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysVal 370
 Db 289 ATGGAACAGATAGCTTAATAAATAATATGATGTCGCGTGAAGAACGACGCTGCGC----- 236
 QY 371 GlyIlyValValProGlyAlaGlyIlyTyraSp 381
 Db 235 GCTTCTAAACTTCAGGTGCTGGTGGTGGCAT 203
 RESULT 32
 AE016745 300242 bp DNA linear BCT 01-JAN-2003
 LOCUS AE016745
 DEFINITION Staphylococcus epidermidis ATCC 12228, section 2 of 9 of the
 complete genome
 ACCESSION AE016745 AE015929
 VERSION AE016745.1 GI:27314765
 KEYWORDS
 SOURCE Staphylococcus epidermidis ATCC 12228
 ORGANISM Staphylococcus epidermidis ATCC 12228
 REFERENCE 1
 AUTHORS Zhang, Y., Ren, S., Li, H., Fu, G., Lu, L., Lu, G., Jia, J., Tu, Y.,
 Qin, Z., Chen, Z., and Wen, Y.
 TITLE Direct Submission

JOURNAL Submitted (05-NOV-2002) Chinese National Human Genome Center at
Shanghai, 250 Bi Bo Road, Shanghai 201203, China

FEATURES

source

1. 300242
/organism="Staphylococcus epidermidis ATCC 12228"

/mol_type="genomic DNA"

/strain="ATCC 12228"

/db_xref="ATCC:12228"

/db_xref="taxon:176280"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735

/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

127. . 735
/locus_tag="SE0305"

```
/product="elongation factor EF-Tu"
/protin_id="AA003909.1"
/db_xref="GI:27314773"
/translation="MAKEKEDSKHEANIGTIGHVDHGTTLTAATATVANKDPVA
OSVMDTNAPEEKERGITTNTAIEYQDCKRYAHYDCGHALDYKMTTGAONDCG
ILVNSAADPMOTREHIIILSRNVGFALVFLANKVDVDEBELVEMERDLISE
YDPGSDVPIVAGSALKALEGAEYEQKLDLMOADDYIIPPERSDRPFMMPEVDY
FSITGRGATVATGVERGOIKVEBEVILQMHETSQTVAQVEMERKLDLVAAGNIG
ALRGVAREDOGVLAAGSITPHTKEAYVLSKDEGKRPFPFNTRPOFYER
TDTGVNVLPEGTEWMPGDVEMVELIAPAIEDGRFSIRSGRTVSGVATEI
FE"
complement(13620..14789)
/locus_tag="SE0313"
complement(13620..14789)
/locus_tag="SE0313"
/codon_start=1
/cran1_table=1
/product="N-acyl-L-amino acid amidohydrolase"
/protin_id="AA003910.1"
/db_xref="GI:27314774"
/translation="MEDMFOLASNKEKRWOLRRYLHOYBELSFEEKRTHDPVNOIS
OLACTIETPVNGRIKATFKGSDNSNGPTIALADPALPVQSLNDVPRSRKGMHA
CCHDGTAILLGVAEIVHEHRLKKNVFIPOKEEIMPSSQEMIDGCIQNDYKI
YGTIWSGIPSGITISRGALVNSPDESVITTYGKGHAKPHETIDPVIWAEFIS
AOKIISRTIDPVKEAVLTFGMIQAGSTSDVDPFAKCTGTFTPTKQSHVQNDK
LIQGLSLNDITYELEIKGYLPHNHQOSYDVVQOANDLHREFESDIAMIGDPS
HLKVRPGAPFLTGCGNKDKGITAPHNHFPIDESSLKVASEFLKILEINVE"
15447..15872
/locus_tag="SE0314"
15447..15872
/locus_tag="SE0314"
/codon_start=1
/cran1_table=1
/product="unknown"
/protin_id="AA003911.1"
/db_xref="GI:27314775"
/translation="WVPENTDIANSFENNMTDLNCEVDEGFVPTTLFEYKDSII
IGAVDIRHQLNDKSNIGHVGVAKSRGKVATILIEKELDELKTLINVEVLMTC
NELNFASTQVMKCGQYIESYIKNGKRVHRVHLPNTEK"
16124..17074
/locus_tag="SE0315"
16124..17074

Alignment Scores:
Pred. No.: 0.0036 Length: 300242
Score: 181.50 Matches: 99
Percent Similarity: 36.74% Conservative: 52
Best Local Similarity: 24.09% Mismatches: 131
Query Match: 8.14% Indels: 129
Gaps: 18

US-10-069-062-7 (1-432) x AB016745 (1-300242)
QY 7 AIAAPROGLYLYALAPheleuAAGLYTYrleuValleuGluProIleTYrAspAla 26
Db 60562 GCCCCCGGAAACCTTATATTCAGGCGAGTATGACGTAACCAACGAGATATTAATCT 60621
QY 27 TYrValThraIeuserSerarGmetHisAlaValIle----- 39
Db 60622 ATCTTATTGCGGTAATATGCTTTGTAAACGCGACATATGAGCGCTCAAAATTAAGTTGAA 60681
QY 40 -----ThrProlyseGlyThrSerleuylGluSer 49
Db 60682 GGTAGTATTCATCCAAACATTATATGAAACGATTAATTTGACCGTATATGAAGT 60741
QY 50 ArgIlelylIleSerSerProGlnPheAlaenglyGluTrpGluTyHisIleSerSer 69
Db 60742 AGAATGGAATCTCAAGATCTTAAGCTGTACGAA----- 60777
QY 70 AenThrgluTyProArggluValGlnSerArgIleAenProPheleuGluAlaThrIle 89
Db 60778 -----CTGAATATATGTTGTC 60792
QY 90 PheIleValleuAlaTYrIleGlnProThrgluAlaPheAspLeuGluIleIleIleTYr 109
```

```
Db 60793 ACAGCTAATAGAGTGTTCGAAACAGTATGCGCAGCTTGCAATATGAAATTA----- 60843
QY 110 SerAapProGlyTYrHisSerGlnIleuSerThrgluThrylTyHisSerSeranglyGlu 129
Db 60844 -----AACGACTTCAATTTAAACATGATATGATTAAGTACGATTAACCTTGGT----- 60891
QY 130 LysThrPheleuTyHisSerArgAlaIleThrgluValGluLysThrGlyleuGlySer 149
Db 60892 -----CAGAATGACGATTAAGTTCA 60912
QY 150 SerAlaGly---LeuValSerValValAlaThrSerleuLeuSerHisPheIleProAen 168
Db 60913 AGCGCGCGTGTTCATTCGTCGTT-----AAAGCTTGAATGAATTCATGCT----- 60963
QY 169 ValIleSerThraLysAspIleleuHisAenValAlaGlnIleAlaHisCysTYrAla 188
Db 60964 ---TTGAATTAATCAACCTTATATTAATTAATGCTGTAATGCAATATGAAATTA 61020
QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTYr---GlyleuIleVal 207
Db 61021 CAAGCTTAAGTTCATGTCG---GATATGCGGTTAGTGTCTACAGTGTGCTTGCA 61077
QY 208 TYrArgArgPheGlnProAlaIleuIleasnApe---ValPheGlnValleuGluSerAsp 226
Db 61078 TATAGTACGTTTGAC-----CATGACGCGTGGAACACGAAATGAGAA----- 61122
QY 227 ProGluLysPheProThrgluLeuLysLysleuIleGluSerAsnTrpGluLysHis 246
Db 61123 -----ACATCGGTGAATGATGTTTGGAAAAAATTTGGCCAGGCTTAACAT 61167
QY 247 ---GluArgCysThrLeuProTYrGlyIleLysleuLeuMetGlyAspValLysGlyGly 265
Db 61168 ATCGAATTCATTACAAAGCTCCGAAATATGAAAGCTTATATGATGATGACGCTGCTCA 61227
QY 266 SerGlnThrProLysLeuValSerArgValleuGlnTrpLysGlyLysProGluGlu 285
Db 61228 CTTCTCTCCACACTTATGATGATGAGTCAACGCTTAATA----- 61269
QY 286 SerSerValValTYrAspGlnleuAsnSerAlaenLeuGlnPheMetLysGluLeuArg 305
Db 61269 ----- 61269
QY 306 GluMetArgLysLysTYrAspSerAspProGluThrTYrIleLysGluLeuAspHisSer 325
Db 61270 -----TCAGATCAAGTTTATGATGATTTTATGATCAATCT 61308
QY 326 -----ValGluProleuThrValAlaIleLys-----AsnIle----- 336
Db 61309 CATGCTGTGTAGAAAGTTTATATCCAGCTTTTAAACATAATATATCAAGGTGTCA 61368
QY 337 -----ArgLysGlyleuGlnAlaIleThrgluLysSerGlyValPro 350
Db 61369 AAGATGATACGTATTAACAGACGATTAATCAACTATATGATGATTAACGACATCAAGTTGAA 61428
QY 351 IleGluProAspValGlnThrgluLeuLeuAspArgCysGlnGluIleProGlyCysVal 370
Db 61429 ATTGAACAGATTAAGCTTAATAAAATTTATGTATGTCGGTGGAAGACGATGCGC----- 61482
QY 371 GlyGlyValValProGlyAlaGlyGlyTYrAsp 381
Db 61483 GCTTCTAAACTTACGATGCTGATGATGCGCAT 61515

RESULT 33
AP004824 290150 bp DNA linear BCT 20-DEC-2002
LOCUS AP004824
DEFINITION Staphylococcus aureus subsp. aureus MW2 DNA, complete genome,
strain:MW2, section 3/10.
ACCESSION AP004824 BA000033
VERSION AP004824.1 GI:21203693
KEYWORDS
SOURCE Staphylococcus aureus subsp. aureus MW2
ORGANISM Staphylococcus aureus subsp. aureus MW2
```

REFERENCE
1 Bacteria; Firmicutes; Bacillales; Staphylococcus.
AUTHORS
Baba,T., Tkeuchi,F., Kuroda,M., Yuzawa,H., Aoki,K., Oguchi,A.,
Nagai,Y., Iwama,N., Asano,K., Naimi,T., Kuroda,H., Cui,L.,
Yamamoto,K. and Hiramatsu,K.
TITLE
Genome and virulence determinants of high virulence
community-acquired MRSA
JOURNAL
Lancet 359 (9320), 1819-1827 (2002)
MEDLINE
22040717
PubMed
12044378
2 (bases 1 to 290150)
REFERENCE
Director-General, Biotechnology Center, Aoki,K., Oguchi,A.,
Nagai,Y., Asano,K., Iwama,N., Baba,T., Kuroda,M., Hiramatsu,K. and
Kikuchi,H.
TITLE
Direct Submission
JOURNAL
Submitted (06-MAR-2002) Director-General, Biotechnology Center,
National Institute of Technology and Evaluation, Biotechnology
Center; 2Chome 49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:bioente.go.jp, URL:http://www.bio.nite.go.jp/
Tel:81-3-3481-1933, Fax:81-3-3481-8424)
location/Qualifiers
FEATURES
source
1.290150
/organism="Staphylococcus aureus subsp. aureus MW2"
/mol_type="genomic DNA"
/strain="MW2"
/db_xref="taxon:196620"
345..1745
/gene="prop"
345..1745
/gene="prop"
/note="ORFID:MM0528
proline/betaine transporter homologue"
/codon_start=1
/transl_table=11
/protein_id="BAB94393.1"
/db_xref="GI:21203694"
/translation="MDPNKENIMVDKAKKTVATGIGNAMWDFGVAYATTAVI
GANFSPVENADIRKMLTFALAIAPLIRPIGVVFGIIGDKYGRKVLSTILMAF
STLTGLPSYDIOGLMAPLILALAVLQSFSGEYAGAMTVABSSPDKRNISGS
GLEIGTSGTASIMIAVLTPELTDEQMASFGMRIPILGLGLGILGRKLEES
PVEENVATQPERKDNINFLQIIRFYITDIFVCVAVVFFVNTVMTATVLTLEVI
KLDATTSVLTICVMAIMIPLAMPGLADKIGEKVPLITGTGLTFSTIANMLHS
OSFVVIVIGIFILGFSTYEATMPSLPTMPFYSRLITSTVENISVIFGCTPL
VATLVTKGSDPLAPAYVLTALISVIGPLVITPLHLSTAGSLKSGSYPNVNDQDRAY
AARPKRALMVKERKN"
2080..2289
/gene="MW0529"
2080..2289
/gene="MW0529"
/note="ORFID:MM0529"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB94394.1"
/db_xref="GI:21203695"
/translation="MSVIAIATSPFLYIHVTSYLIKNNVLYQYVILGKFDPRNQV
YLIKILHYMLTICLIKFTLSRGY"
2289..3665
/gene="vraA"
2289..3665
/gene="vraA"
/note="ORFID:MM0530
hypothetical protein, similar to long chain fatty acid CoA
ligase"
/codon_start=1
/transl_table=11
/protein_id="BAB94395.1"
/db_xref="GI:21203696"
/translation="NNVILEQLKTHTONKNDIALHIDDETIYSQLNARISAVESL
OKSLNPVVAINKSPVQSTICVLAHLRHAKVPMMEGKQSTIHQLEKYGIVKI
GDTGLQNDISEWFIIDSTQLOHYENLHLGFTSGTGLPKAVYRDEDSWASVEVM
LMKNNENAIAPGLSHSLTLVALPALSGRFTIQOTFHPKPLNQCHKLSYVVA
MFLVPTMIXSLILVNNHETIQSFSSGDKLHSSIPKIKKNQNDINLIEFQTSRTS

gene
367..4806
/gene="vraB"
3667..4806
/gene="vraB"
/note="ORFID:MM0531"
/codon_start=1
/transl_table=11
/product="acetyl-CoA C-acetyltransferase"
/protein_id="BAB94396.1"
/db_xref="GI:21203697"
/translation="MNQAVYAKRTAFKGYGTLKHLBPQLKPLRQHPKRYPV
ISKIDVYLVNGVNGMIRAKLLBALKKSIIPGVTIDROGSGLESGVQACMTQIA
GAKYVINGVSTSRAPKIKRPISVETALPEYERASFPBSSDSMTGLAENVA
KNYDVSRELQDEFAVRSQTLAENYKNNSIQELIPVKGSIPTNDSLSKHIPDN
RFRFPVYIKGTVTAANSCKMKDGAVALIMKEDMAVEBPHGLPLFDGVTVGDSN
FGIGPVPASISLAKRNQLTENIEVIRINEAFSAQVAVACQALNISTQNLNMGAL
ASGHPYASGAQLVTRLFVMPDKETIMSMGIGGLGNALPTRF"
4781..5146
/gene="vraC"
4781..5146
/gene="vraC"
/note="ORFID:MM0532"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB94397.1"
/db_xref="GI:21203698"
/translation="MQHYLDSNORLNTSPSKDSVAAYQCFNQPYRKEVPLMCASL
WPKPDLFKKYNSELILTKSAINQOKLEVDTIYVGHLEIDRCQTRNITRYTMAULT
TKNDQHVITVQTFIKAMK"
5149..5418
/gene="MW0533"
5149..5418
/gene="MW0533"
/note="ORFID:MM0533"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB94398.1"
/db_xref="GI:21203699"
/translation="MKFNEIMINEYALVNDNPIHNRIVPSQVLSQWMLMMSLETN
OQQIVYKPILENINIEPIEQHEHIIAINDGEIKIKISLSTR"
complement (5619..5786)
/gene="MW0534"
complement (5619..5786)
/gene="MW0534"
/note="ORFID:MM0534"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB94399.1"
/db_xref="GI:21203700"
/translation="MIYRQYHNEGAPVYEIITKTFQVHSIKCDSFSQTEIFKLLSL
LQDDIDHMKVS"
complement (6293..7123)
/gene="MW0535"
complement (6293..7123)
/gene="MW0535"
/note="ORFID:MM0535
hypothetical protein, similar to phosphomethylpyrimidine
kinase, chid homologue"
/codon_start=1
/transl_table=11
/protein_id="BAB94400.1"
/db_xref="GI:21203701"
/translation="MAKLVLTIASDTSAGAGMADLTROELDTYGVNLTATVTM
DQVTSHTDTPLEMPVFEKQLTALSIGBDATKTMGLTGEETIKRAGEVTEASNAQF
VVDPAWVCAGEBEVIANPNTETAMITRYLLPKAVVTPNLFPAQQLSGLGKLNIEIDKQ
AAITLIFDKAGQVIVIKKGKALDQDSYDLVYDQGFYQYLTIDMFOOSVNHGAGCTFAA

LOCUS DEFINITION	3074 bp DNA linear	BCT 29-AUG-2000
AF290095	Enterococcus faecium mevalonate kinase (mwak1), mevalonate diphosphate decarboxylase (mwad), and phosphomevalonate kinase (mwak2) genes, complete cds.	
ACCESSION VERSION	AF290095	
SOURCE KEYWORDS	AF290095.1 GI:9937392	
ORGANISM	Enterococcus faecium	
REFERENCE	Enterococcus faecium Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.	
AUTHORS	1 (bases 1 to 3074) Wilding,E.I., Brown,J.R., Bryant,A.P., Chalke,A.F., Holmes,D.J., Ingraham,K.A., Iordanescu,S., So,C.Y., Rosenberg,M. and Gwynn,M.N. Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in gram-positive cocci	
JOURNAL MEDLINE	J. Bacteriol. 182 (15), 4319-4327 (2000)	
REFERENCE	20353468	
AUTHORS	10894743	
TITLE	2 (bases 1 to 3074) Wilding,E.I., Brown,J.R., van Horn,S., Mathie,T., Warren,R. and Gwynn,M.N.	
JOURNAL	Direct Submission	
FEATURES	Submitted (26-JUL-2000) Department of Microbiology, Smithkline Beecham, 1250 S. Collegeville Road, Collegeville, PA 19426, USA	
source	location/Qualifiers	
	1..3074	
	/organism="Enterococcus faecium"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:1352"	
	62..1006	
	/gene="mwak1"	
	62..1006	
	/gene="mwak1"	
	/note="mwak1"	
	/codon_start=1	
	/transl_table=1	
	/product="mevalonate kinase"	
	/protein_id="AAG02445.1"	
	/db_xref="GI:9937393"	
	/translation="MANYGQESSGKIIIMEHVAVYGEPAIAPPATVATVATLEELDAMDQVLSYYSNGNLAAEPAPAKKIKKLFHKKQNDIQNDLITESTIPARGCNGSSAAVAIVATVAFDYIALFPLSRISLILENVQSLKIAHGNSGIDAATSLQTYPTTGSHHPDYSLINIDAFILVADYIGKQREAVKVAHLETPQPHGTQMIOKLGYLTKQAKOALIIENSPTLIQTMDESOSLIEKLTISDNFLIIOPTKQMGALGAKLTGGGRGGCMILAQRTKTAQELISQLEDAGAAETWIGLGVHTYV"	
	999..1976	
	/gene="mwad"	
	999..1976	
	/gene="mwad"	
	/note="mwad"	
	/codon_start=1	
	/transl_table=1	
	/product="mevalonate diphosphate decarboxylase"	
	/protein_id="AAG02446.1"	
	/db_xref="GI:9937394"	
	/translation="MEFGKARAYTNIALIKYMGKKNEELIPMNSISLITADAYETETEVIRSDSYMVEFYIDGTLODEKATKRYSGPLDIPREAGSLKASVTSQNFVYTAAGLASSSGAALAGACNTILKVLGDLDSLRFARRSGSAGRSITGGFEWKEGHDLYQSYAKRVPDSFEDDLANVFLINDQKEVSRNMRRTVETSNFYQGMLSVEEDLYQ LKQALIKTQDPLLGETMERNGLIKMGITLAAQPTVYSPNSLKAMDAVQLRKQGIPI CYFTMDAGPNVKAIVENSHLSEVQETFKLPSKQVITAHAGPGLALIE"	
	1989..3074	
	/gene="mwak2"	
	1989..3074	
	/gene="mwak2"	
	/note="mwak2"	
	/codon_start=1	
	/transl_table=1	
	/product="phosphomevalonate kinase"	
	/protein_id="AAG02447.1"	

DB	BASE COUNT	953 a	621 c	662 g	838 t
ORIGIN					
Alignment Scores:					
Score: No.:	1.97e-05	Length: 3074			
Percent Similarity:	177.50	Matches: 97			
Best Local Similarity:	29.55%	Conservative: 62			
Query Match:	24.13%	Mismatches: 144			
DB:	7.96%	Indels: 99			
	1	Gaps: 18			
US-10-069-062-7 (1-432) x AF290095 (1-3074)					
QY	6	SeRLaPrroGlyLysAlaPheLeuAlaGlyGlyTyrLeuValLeuGluProIleTyrAsp	25		
DB	2001	TCCTCACACGACCAACTCTATATTGCGGAGAAATATCAGTGTGTGAACAGCCATCCA	2060		
QY	26	AlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLysGlyThrSer	45		
DB	2061	GCAATTATCGGTGACAGTCATCAATTCGTGACACTGTA	2102		
QY	46	LeuLysGluSerArgIleLysIleSerSerProGlnPheAla	59		
DB	2103	GAATCCGACAGAAAAGTCGAAAGTATCAACTGCACAAATATAGCGAGTCCT	2156		
QY	60	AsnGlyLysTrpGluTyrHisIleSerSerAsnThrGluLys	73		
DB	2157	GTAAGTCGGACACAGCAACGCGAAGTATGATTTCATAT	2198		
QY	74	ProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeu	93		
DB	2199	CGGAGAAATTCCTTT	2225		
QY	94	AlaTyrIleGlnProThrGluAlaPheAspLeu	111		
DB	2226	GCTCTATTCGCTTGCATGAAAGTATGACACAAAGAAAAAATTCCTTTA	2276		
QY	112	ProGlyTyrHisSerGlnIleAspThrGluThrIleThrSerSerAsnGlyGluLysThr	131		
DB	2277	TCCTTATATGATCTGAAAGTACAAAGTAGTAAACAGTTCAAATGGC	2324		
QY	132	PheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAla	151		
DB	2325	CGGAAATATGGTTTGGGTTACG	2348		
QY	152	GlyLeuValSerValAlaIleThrSerLeuLeuSerHisPheIleProAsnValIleSer	171		
DB	2349	GGTCCCGTACCGTTGACACGCTCAAGACCTTCGAATCTTTTTCAC	2402		
QY	172	ThrAsnLysAspIleLeuHisAsnValAlaGlnIleLeuHisCysTyrAlaGlnLysLys	191		
DB	2403	TTATCTCAGTTGGAGATTTTTCACATATGGACACTATGCAATTTAGCAATTCAGATAT	2462		
QY	192	IleGlySerGlyPheAspValAlaIleThrAlaIleTyr	210		
DB	2463	GGTTCCTCGCGGCGACATCGCTGCTAGAGTATGATGGTGGCTGATCGCTTCTCAAC	2519		
QY	211	PheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGluLysPhe	230		
DB	2520	TTTCAGCACCATCTTGCTTCACAGAA	2552		
QY	231	ProThrGluLeuLysLysLeuIleGluSerAsnTrpGluLys	249		
DB	2553	CAGCAATTCATCAATGAGATTACTTGGCCCTGGATGGCCAGGTCTATGCATTTAGACCATTTG	2612		


```

Db      940 TCTAAACATCAGCGCGTGGTGGAGACTGT---GGTATTCATATTCATTAAGAT 996
QY      392 ValGlyAsnPhelys 396
Db      997 GTGATTAAGAAAA 1011

RESULT 36
AF290087 3145 bp DNA linear BCF 29-AUG-2000
LOCUS     Staphylococcus aureus mevalonate kinase (mwak1), mevalonate
DEFINITION diphosphate decarboxylase (mwad), and phosphomevalonate kinase
            (mwak2) genes, complete cds.
ACCESSION AF290087
VERSION    AF290087.1 GI:9937363
KEYWORDS
SOURCE     Staphylococcus aureus
ORGANISM   Staphylococcus aureus
REFERENCE  1 (bases 1 to 3145)
            Bacteria; Firmicutes; Bacillales; Staphylococcus.
AUTHORS    Wilding,E.I., Brown,J.R., Bryant,A.P., Chalker,A.F., Holmes,D.J.,
            Ingraham,K.A., Iordanescu,S., So,C.Y., Rosenberg,M. and Gwynn,M.N.
            Identification, evolution, and essentiality of the mevalonate
            pathway for isopentenyl diphosphate biosynthesis in gram-positive
            cocci
JOURNAL    J. Bacteriol. 182 (15), 4319-4327 (2000)
MEDLINE    20353468
PUBMED     10894743
REFERENCE  2 (bases 1 to 3145)
            Wilding,E.I., Brown,J.R., van Horn,S., Mathie,T., Warren,R. and
            Gwynn,M.N.
TITLE      Direct Submission
JOURNAL    Submitted (26-JUL-2000) Department of Microbiology, Smithkline
            Beecham, 1250 S. Collegeville Road, Collegeville, PA 19426, USA
FEATURES
Source     1. 3145
            /organism="Staphylococcus aureus"
            /mol_type="genomic DNA"
            /db_xref="taxon:1280"
            48..968
            /gene="mwak1"
            48..968
            /gene="mwak1"
            /note="mwak1"
            /codon_start=1
            /transl_table=11
            /product="mevalonate kinase"
            /protein_id="AA02424.1"
            /db_xref="GI:9937364"
            /translation="MTRKGYESTGKIIIGEHAVTGEBAIAPVFNAGKIKYLIAL
            /ESGNYSIKSDVDGMLYADPHLKSIVNFELNMTBPLAVTIGTNPSPGSGS
            AAVAVAPVRASVDPLKSLTKELIRANABOLARIGSGASRSIYGGPAWBGDQ
            QATTLTSLSDGVMVYIDTVKSGTQADVHKLCEDPQYMSHVHIGLTVRASDV
            IEHNFEALADIFNECHADIKALVSHDKLEQMLKIGKENGALAGKLTGARGGSMIL
            LAQDLPARKIVAVAKAGAAHTMIENLGG"
            973..1956
            /gene="mwad"
            973..1956
            /gene="mwad"
            /note="mwad"
            /codon_start=1
            /transl_table=11
            /product="mevalonate diphosphate decarboxylase"
            /protein_id="AA02425.1"
            /db_xref="GI:9937365"
            /translation="MIRSGARATNTALIKWGKDEALIPNNISIVTLKFEYTE
            TKYTFNDQLQDFQNGEAVSGELEKISKYMDIYVRAGIDVVAIEBNDKVPAA
            GLASSASVAALAAACNOALDLOSDKLSRLARIGSGASRSIYGGPAWBGDQ
            TSAVAPLESNHPEDDLAMTFVYNHSGKVPKSGSTPTNTSPRYQMDYDDELIA
            EAKAIDODKPKFLGVEIENGIRMAHTNGSTPPPTIYQESYDVAALVHECREKGY
            PCYTFMDAGNVAVILVEKKNKQIIDLTLQFNNOIIDSIIATGIEIIE"
            1969..3045
            /gene="mwak2"

```

```

CDS      1969..3045
            /gene="mwak2"
            /note="mwak2"
            /codon_start=1
            /transl_table=11
            /product="phosphomevalonate kinase"
            /protein_id="AA02426.1"
            /db_xref="GI:9937366"
            /translation="MIQVAPGKLYAGEAVTEPRYSVLIALDRFVATIEATQY
            KETISKALHNHPVPRPDSIVISDPHAAKOLNVVTAIEIPROVAKSCDIAMKHF
            HTTISNLDNSGKTYGLSSAAVLVSATKVLNERPDMKLSLTYKLAIVANMLQS
            LSSGDIASVSGWMLAYTFDHWKHOIEBTVEVLIKWPGHIEPLQAPENME
            VIGMTGSPASSPHVSEVYKRLSDSPFEGDLSEHSRCVELIHAFTKNNIKGVQKM
            VNQRITIIQRMDEKATVIDIETKLYLCIDIAKRYGASITSGAGGDCGIIITNKDVD
            KERIDEMTKHGIKPKLENIYHGO"

BASE COUNT 1105 a 444 c 678 g 918 t
ORIGIN
Alignment Scores:
Pred. No.: 4,95e-05 Length: 3145
Score: 173.00 Matches: 97
Percent Similarity: 36.24% Conservative: 57
Best Local Similarity: 22.82% Mismatches: 143
Query Match: 7.76% Indels: 128
DB: 1 Gaps: 20

US-10-069-062-7 (1-432) x AF290087 (1-3145)
QY      7 AlaProGlyLysAlaPheLeuAlaGlyIYrLeuValLeuGluProIleYrAspAla 26
Db      1984 GCACCCGGAACCTTATATGCTGAGAAATGCTGTAAACGACAGATATAATCT 2043
QY      27 TyrValThrAlaLeuSerSerArgMetHisAlaValIleThrPro-----Lys 42
Db      2044 GTAATTATGCGTTAGATGTTTGTACTGCTACTATTGAAGACCAACCAATATA 2103
QY      43 GlyThr-----SerLeuysGluSer 49
Db      2104 GGTACCATTCATCAAAAGCATTTACATCAATACCAAGTACATTTAGTAGAGTAAGAT 2163
QY      50 ArgIleLysIleSerSerProGlnPheAlaAsnGlyGluTrpGluTrhIleSerSer 69
Db      2164 AGTATTGTCATTCACATCCACATCGACGCAAAACA----- 2199
QY      70 AsnThrGluLysProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThrIle 89
Db      2200 -----TTAAATATATGCGTC 2214
QY      90 PheIleValLeuAlaTyrlleGlnProThrGluAlaPheAspLeuGluIleIleIleYr 109
Db      2215 ACAGCTATTGAATAATTGTAACAAATACGCAAAAGTTGCGATATACGATG----- 2265
QY      110 SerAspProGlyTyrlhIleSerGlnGluAspThrGluTrhLysThrSerSerAsnGlyGlu 129
Db      2266 -----AACCATTTATCTGACTATGATAGTAATTTA---GATGATTCAAATGCT--- 2313
QY      130 LysThrPheLeuTyrlhIleSerArgAlaIleThrGluValGluLysThrGlyLeuGlySer 149
Db      2314 -----CARAAATATGATTAAGGTTCA 2334
QY      150 SerAlaGly---LeuValSerValAlaIleThrSerLeuLeuSerHisPheIleProAsn 168
Db      2335 AGTGCAGCAGTACTTGTCTGAGTTAA-----AAGTATTAATGAATATATGATATG 2388
QY      169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
Db      2389 AAGTTATCTAAATTTA-----TACATTTATTAACCTGACGATGTCACCAATATGAAGTTA 2442
QY      189 GlnLysLysIleGlySerGlyPheAspValAlaIleThrAlaIleGlyr---GlyLeuIleVal 207
Db      2443 CAAAGTTTAAAGTTCAATCGGA---GATATTCGTCGAGGTATATATGATGATGCTTAAGC 2499
QY      208 TyrArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspPro 227

```

```

Db      2500 TATAGTACTTTGATCATGAATGGGTAAACATCAATGAA----- 2541
Qy      228 GATLysPheProThrGluLeuLysLysLeuLleGluSerAsnTrpGluLysHis--- 246
Db      2542 -----GATACACGGTTGAGAGAGTTTATATCAAAAAGCTGGCGCTGANTGCATC 2592
Qy      247 GUAATGCTythrLeuProTyrGlyLleLysLeuLeuMetGlyAspValLysGlyGlySer 266
Db      2593 GAACCATTCAGACACCTGAAAATATGAGAAGTACTTATCGCTTGCACCTGCACCGCG 2652
Qy      267 GUAATGCTythrLeuValSerArgValLleuGlnTrpLysGlyLysProGluGluSer 286
Db      2653 TCATCACCCACACTTGTGTAGCCAGAGTAAACGTTGAAA----- 2691
Qy      287 SerValValTyrAspGlnLeuLysSerAlaLeuLeuGlnPheMetLysGluLeuArgL 306
Db      2691 ----- 2691
Qy      307 MetArgGluLysTyrAspSerAspProGluThrTyr-----LleLysGluLeuAsp 323
Db      2692 -----TCAGATCCTTCATTTACGGTACCTTCTTGAAGATTCACAT 2733
Qy      324 HisSerValGluProLeuThrValAlaIleLys---AsnIleArgLysGlyLeuGlnAla 342
Db      2734 CGTTGTTGTAAGAGCTTATTCATGCTTTTAAACAAATACATTAAGGTGCAAAAG 2793
Qy      343 LeuThrGlnLysSerGluValProLleGlu-----ProAspVal 355
Db      2794 ATGGGCGCGTACAGATCGTACAAATATTCACCTAGATTAAGAAAGTACAGTTATTA 2853
Qy      356 GlnThrGln-----LeuLeuAspArgCysGlnGluLleProGlyCysValGly 371
Db      2854 GAACCTGAAACCTAAATATTTGTGTGATATTCGCTGAAAGATATCACCGT-----GCA 2907
Qy      372 GlyValValProGlyValGlyLysTyrAspAlaIleAlaValLeuValLeuGlnGln 391
Db      2908 TCTAAACATCAGCGCTGTGTGTGAGACGT---GGTATTACATATTCATTAAGAT 2964
Qy      392 ValGlyAsnPhelLys 396
Db      2965 GTAGATTAAGAAAAA 2979

RESULT 37
AP003131 299050 bp DNA linear BCT 11-JAN-2003
LOCUS Staphylococcus aureus subsp. aureus N315 genomic DNA, complete
DEFINITION AP003131 BA000018
ACCESSION AP003131.2 GI:14349173
VERSION
KEYWORDS
SOURCE
ORGANISM
Staphylococcus aureus subsp. aureus N315
Staphylococcus aureus subsp. aureus N315
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE
1
AUTHORS
Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,
Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Iino,J., Ito,T., Kanamori,M.,
Matsunaru,H., Maruyama,A., Murakami,H., Hasegawa,A.,
Mizutani-Ui,Y., Takahashi,N.K., Sawano,T., Inoue,R., Kato,C.,
Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S., Yabuzaki,J.,
Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C.,
Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiramatsu,K.
Whole genome sequencing of methicillin-resistant Staphylococcus
aureus
JOURNAL Lancet 357 (9264), 1225-1240 (2001)
MEDLINE 21311952
PUBMED 11418146
REFERENCE 2 (bases 1 to 299050)
AUTHORS Director-General, Biotechnology Center, Aoki,K., Oguchi,A.,
Hasegawa,A., Nagai,Y., Kuroda,M., Hiramatsu,K. and Kikuchi,H.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-2001) Director-General, Biotechnology Center,
National Institute of Technology and Evaluation, Biotechnology

```

Center, 2Chome 49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
 (E-mail:bioente.go.jp, URL: http://www.bio.nte.go.jp/,
 Tel:81-3-3481-1933, Fax:81-3-3481-8424)
 On Jun 12, 2001 this sequence version replaced gi:13700446.
 Location/Qualifiers
 1. 299050

COMMENT FEATURES source

```

/organism="Staphylococcus aureus subsp. aureus N315"
/mol_type="Genomic DNA"
/strain="N315"
/sub_species="aureus N315"
/db_xref="taxon:158879"
248..1324
/gene="11ve"
248..1324
/gene="11ve"
/note="ORFID:SA0512
branched-chain amino acid aminotransferase homologue"
/codon_start=1
/transl_table=11
/protein_id="BAB41743.1"
/db_xref="GI:13700446"
/translation="MSQAVKRRERETLKQKRWTSQSGRGKFTDYMSTYDADKGMH
DKTVPGPIETSPAQGVHYGQSVFBLKAYKKGDEVALFRPENFRLLNSLARLE
MPQVDEALLEGKQLYDIREDWIPEBGQSLYLRPFVATGALGVGASHQYKLI1
LSPGAYYGGETLKPKLYDEVAVRAGGVAKVAGVAAASLLAOTNANKGYDQ
VLAMDGVOKYIEEVGSNNIPFVENKVTPELNGSILPGITRKSIIELANLGYEVE
RRVSIDELPESSVDKELTEVPGSTAAVISPVGLTRERDEIYINNRETELQKLY
DYVTGQNGTLEDKGRVVPVKY"
1574..2257
/gene="SA0513"
1574..2257
/gene="SA0513"
/note="ORFID:SA0513"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB41744.1"
/db_xref="GI:13700447"
/translation="MEMTLPDKDGTLEFDRSMWKIGVRPFVQSLETFPPYNNKAALR
OLGVTKESIDPKSVSGSLQOIIOAFRDVYGDPTDWSKSTQKLVDERIPETNWE
GVKALIDILKARGYGLVTSDTKGVQPLATNNAISLPDIIISTEDADAVEKPKVY
LSPLEEQYNDPQKVAIVGDYANDMKTAASNMLGMAIGVLGIAIKETLRADIIINS
AADIIEALN"
complement(12373..3035)
/gene="SA0514"
complement(12373..3035)
/note="ORFID:SA0514
hypothetical protein, similar to deoxyurine kinase"
/codon_start=1
/transl_table=11
/protein_id="BAB41745.1"
/db_xref="GI:13700448"
/translation="MNNYGIIPONAIITAGTGVGKSTLQALADKLNFKTSFENVEH
NPYDKFSDPERWSFHLQYFLAERFQGRMEYGGGFVQDSIYEDVDFPAMME
BEGMSKSEDFEFTYSDFLNAMVMPFPPKPMVILENVEVIDRITIERGEMINTD
PEYVAKLIPKRYDDMINSFNACPVYRININEYDIHKDESLNPMIDKARIITQTRQVD
TR"
complement(3028..3645)
/gene="SA0515"
complement(3028..3645)
/note="ORFID:SA0515
hypothetical protein, similar to deoxyurine kinase"
/codon_start=1
/transl_table=11
/protein_id="BAB41746.1"
/db_xref="GI:13700449"
/translation="MKRPFATISRPVGVGKSLAHKLSQTLDFPEKEIITENPPLSD
FYBISKWSQTEFPLCNRPKOPQDTQLNGQVSDYHAIKNNKIFAKNTLSPEVFK
FSKIVLITLDMIMPMNIIFLADLDVLKRIARUNRSPHGIEDVYLKIKQVREY
YESLOSNGSNVVIDTTSIDPLKNEQYEDILHILPMIGDITWE"
3712..4182

```


QY 169 ValIleSerThrAsnLysAspIleLeuHlaSerValAlaGlnIleAlaHisCysTyrAla 188
 Db 44093 AAGTTATCTTAATTTA-----TACATTATTAACCTGACGTGCAATGCAATGAACTTA 44146
 QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyr---GlyLeuIleVal 207
 Db 44147 CAAAGTTTAAGTTTCATGCGGA--GATATTCCTGAGAGGTATATAGTGAATGAGTGAAGG 44203
 QY 208 TTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 44203
 Db 44204 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 44204
 QY 228 GlnLysPheProThrGlnLeuLysLysLeuIleGlnSerAsnTyrGlnLysHis--- 246
 Db 44246 -----GATATTCCTGAGAGGTATATGCAAAACTGCGCTGATGACATG 44236
 QY 247 GlnArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlySer 266
 Db 44297 GAACCATTAACAAGCCTGATAAATATGAAAGTATATGATGATGATGATGATGATGATGATGAT 44356
 QY 267 GlnThrProLysLeuValSerArgValLeuGlnTyrLysGlnLysProGlnLysSer 286
 Db 44357 TCATCCACCACTTTGTTAGCAAGTGAACCTTGAAA----- 44395
 QY 287 SerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGlnLeuArgGln 306
 Db 44395 ----- 44395
 QY 307 MetArgGlnLysTyrAspSerAspProGlnTyrTyr-----IleLysGlnLeuAsp 323
 Db 44396 -----TCAATTCCTTCATTTACGGGACTCTTCAAGAACTTCAAT 44437
 QY 324 HisSerValGlnProLeuThrValAlaIleLys---AsnIleArgLysGlnLeuGlnAla 342
 Db 44438 CCGTGGTTGAAAGACTTATGATGCTTTTAAACAAATTAACATTAAGCTGCAAAAG 44497
 QY 343 LeuThrGlnLysSerGlnValProIleGln-----ProAspVal 355
 Db 44498 ATGGCGCTGACGATCGTCAATTTATTAACGTATGATGATGATGATGATGATGATGATGATGAT 44557
 QY 356 GlnThrGln-----LeuLeuAspArgCysGlnGlnIleProGlnCysValGly 371
 Db 44558 GAAACTGAAACCTTAATATTTGCTGATTTGCTGAAAGATGATGATGATGATGATGATGATGATGAT 44611
 QY 372 GlyValValProGlnAlaGlyLysTyrAspAlaIleAlaValLeuValLeuGlnLysGln 391
 Db 44612 TTTAAACATCTGAGCGCTGCTGAGACCTGT---GGTATTACAATTATTAAGAT 44668
 QY 392 ValGlyAsnPheLys 396
 Db 44669 GTAGATTAAGAAAA 44683
 RESULT 38
 AF003359
 LOCUS Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete
 DEFINITION Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete
 ACCESSION AP003359 BA000017
 VERSION AP003359.2 GI:14246074
 KEYWORDS
 SOURCE
 ORGANISM
 Staphylococcus aureus subsp. aureus Mu50
 Staphylococcus aureus subsp. aureus Mu50
 Bacteria; Firmicutes; Bacillales; Staphylococcus.
 REFERENCE
 1
 AUTHORS
 Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I.,
 Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Imai, J., Ito, T., Kanamori, M.,
 Matsunari, H., Maruyama, A., Murakami, Y., Hosoya, A.,
 Mizutani, U., Y., Takahashi, N., K., Sawano, T., Inoue, R., Kato, C.,
 Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J.,
 Kanehisa, M., Yamashita, A., Oshima, K., Puruya, K., Yoshino, C.,
 Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H., and Hiramatsu, K.
 Whole genome sequencing of methicillin-resistant *Staphylococcus*
aureus

JOURNAL Lancet 357 (9264), 1225-1240 (2001)
 MEDLINE 21311952
 PUBMED 1148146
 REFERENCE 2 (bases 1 to 343590)
 AUTHORS Ohta, T.
 TITLE Direct Submission
 JOURNAL Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College
 of Medical Technology and Nursing, Department of Medical
 Technology, 1-1-1 Ten-no-dai, Tsukuba, Ibaraki 305-8577, Japan
 (E-mail: tohta@tsukuba.ac.jp, Tel: 81-298-53-3454,
 Fax: 81-298-53-3454)
 COMMENT On May 29, 2001 this sequence version replaced gi:13874621.
 FEATURES
 source
 1..343590
 /organism="Staphylococcus aureus subsp. aureus Mu50"
 /mol_type="genomic DNA"
 /strain="Mu50"
 /sub_species="aureus Mu50"
 /db_xref="taxon:158878"
 441..1331
 /gene="SAV0307"
 441..1331
 /gene="SAV0307"
 /note="similar to outer membrane protein precursor"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="BAB56469.1"
 /db_xref="GI:14246075"
 /translation="MNRISKYLAIASISAVATVAPOTNTSTAPAKSSAEVQOQOAS
 IPASQKANKQNIQNAVAVYQNSAEALAYQGYNSAKYQLEKEIKKNGGKALAL
 LDEETVLDNSPYQGVASIHNPPEGEHMYQAKAPVGAKEFLKADKKGVYIY
 ISDDDKKDKLAKATQNKIKOQIIPAKKSHILLKQDKSKESRQMOQDKLWLRG
 DNLDPTDPRKATRESREALTEKKKDPGKKYIIFPNMYSMEATITNNYKASDKA
 KDKRKAIKQFDKTESEK"
 1580..2629
 /gene="SAV0308"
 1580..2629
 /gene="SAV0308"
 /codon_start=1
 /transl_table=1
 /product="conserved hypothetical protein"
 /protein_id="BAB56470.1"
 /db_xref="GI:14246076"
 /translation="MPLAMNIRNRNKKRFGIITGVLMISYLEPLSLGANGLNK
 EGIDKQADAVLKNQVQOSVFNKKDLENKKQATLKQGEIVSNHQDNVL
 VFGVEKSFVPSLIEGKATKQNEVLADETLKNKGRKIGPTLSQSDGLHVGFT
 ESAKYNASPVIFNDATIAKINPRTGKINAIVVVDPTNMKDKLNLBLEVSNDFI
 ENLPGYKQNLINFMISPLFVISAIVIGLFLVYMTDQKSLFGLIAQOGPTNGVLAN
 VVLSQTLIALFGPARGLLTLGVGALFLPAVPKPKPVTLVPAIVLMIVSVLSLE
 SIIIRKIDPLKAIG"
 2642..3319
 /gene="SAV0309"
 2642..3319
 /gene="SAV0309"
 /note="similar to ABC transporter ATP-binding protein"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="BAB56471.1"
 /db_xref="GI:14246077"
 /translation="MLKEENVTKSPKQNNNIEAVQDTEINPKDITLALVPGSGSK
 STPLTMAGALQTPSKSHILLNNODITTMKQKALAKVMSSEIGFLQTLNVPLPTVKQ
 QFTLLKKKNKNSNEDYQQLMSQLGLSLNKLPSISGGQKQVVAIAKALYTPSI
 ILDEPPLADTEVALIEVIRLDAQARRKAKCIVTHDRLLKAYCDRSYHMGQVILN
 LENVTV"
 3527..4558
 /gene="SAV0310"
 3527..4558
 /gene="SAV0310"
 /note="similar to regulatory protein Pfor"
 /codon_start=1
 /transl_table=1

gene /product="hypothetical protein"
/protein_id="BAB56472.1"
/db_xref="GI:14246078"
/translation="MDMLIGTLFLVLVIFTLPTPKANPGMRMGALANAATPLFLV
EAFNKVGEVEVFGIETLEELGDAAGLGVAAGALATALICVSPVALVIAAGGMD
LIPFGPAGMIGVYMKYETKRYPDGVVDLIGSLVILAPLAILAVLPPVNSTLIRIG
DIIOSSTNMPIMGIIIGGIIITVVGTPALSMALTALLGIGVMAIGMAAFSSAF
MNGTIFPHRLKLDKRKSTTAVSIEPLSQMDIYSANPIPIYITNPFQAGALIGLIMSGL
INDAQTAPLPIAGFLWMPGFPHNPMTIYIVGVMAIVGALAYLGSIVFKKPIYTKOD
MNRGAVDA"

gene 4901..6019
/gene="SAV0311"
4901..6019
/gene="SAV0311"
/note="similar to carbohydrate kinase pfkB family"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56473.1"
/db_xref="GI:14246079"

gene /translation="MSDEKETLKRIKDNPEISORELAELGLSPSVAIIISGLIOK
EYVWKAYVLNEDYPIVICGANVDRKRYKHDVAEASNVSTSTRSIGVARNIARN
LGRLEAFVFLSASGDSMEMIKRLSTPPNMLDVQOOFENAGSYYALISKEGDMT
VGLAMVEVDYITPEELIKRSHLKKACCIIVDNLGKEALNPLCAVTHGQILVIT
TVSSRKMKRMPDLAIDWITNKNDETITLANKESTDDLKIAKARNRDNQVNVIV
TNGVELLYRSGEEHIIKSWPNSVKOVTAGDGSFCAVAVYSLNEMSTEDLIAGM
VNAKTIETKRYPRNLDQOQLYHDMEDYKNGKFTKYV"

gene 5994..6917
/gene="SAV0312"
5994..6917
/gene="SAV0312"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB56474.1"
/db_xref="GI:14246080"

gene /translation="MANLQKTYEYREVOQARENNPIVALESTIISHGMPYONVEM
ATTVEKILNNNAIPATTAIIDGKIKIGLESBDEITLTSKDAKVSERDIAEVANKM
CVGATTVATIMMCAAMAGIQEPVVTGIGVGHAEHTMIDADLEELSKTNVTYICAG
AKSIIIDLPTEHVELETKGVPIGYOTNLPAPFTRESGVKLTSVEVTERADIDHLTK
QOLNEGGIVANPIPYEHALSKAYIEAIINBAVEAENQIKGDMATPFLGKIVKEK
TNGKSLANIKRIVENNALGAKIAVAVKLL"

gene 6928..8148
/gene="SAV0313"
6928..8148
/gene="SAV0313"
/codon_start=1
/transl_table=11
/product="probable pyrimidine nucleoside transport
protein"
/protein_id="BAB56475.1"
/db_xref="GI:14246081"

gene /translation="MNLLPALTGIALFALFVAFLESPDRKNDIFKKTLIMTPIOLYIYL
EMANTTIGTITLALGSPFREGILINYSKAGINPFVEDIONKRGTPPLPLAVLLPVLISV
LIGNTYVIAFLPFIIRKIVGALINKITRNGRLSEFALSTAFNGQPEVYLTIKDILPEL
LSAKIYVITATSGMSAVSAMLGSYQMQLPEKRVYAVMLNFSALIIVASVNPKSDSD
SVEIADINKTSTETKSTGVNGKTKGPKKVAFPQWIGDSANDGKIAVVAVMLAFLISLM
EALINMFSGVGNFKOLIGYVAPAPIAFPLMGVAPMSAIVPAGSLMTKLTINFEVMLDPE
KVLVDVSAKRGCIISVYLVSFANRGYGIIVGSIKGISDKQGRKVSFMRLLLGST
LASIISGSLIGVLT"

gene complement(8253..9785)
/gene="SAV0314"
complement(8253..9785)
/gene="SAV0314"
/note="similar to sodium-coupled permease"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB56476.1"
/db_xref="GI:14246082"

CDS /translation="MKVEGCHLNVAVYIYILAMLPICGVETPKRASQSTNPSFTASG
RLPSVVGSGIATTLTSAITFMSTPEKAFLDWMSITAGNIAIVAILPILITFYPPFKR
KIKVISAHEYLEARGPSIRVIGSLIPVYHGRVAIVIYLPILAITVSQMNQYIVAA

[illegible]

Qy 130 LysThrPheLeuYrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySer 149
 Db 317318 -----CATAAATATGATAGTTAGTTCA 317338
 Qy 150 SerAlaGly---LeuValSerValValAlaIleThrSerLeuLeuSerHisPheIleProAsn 168
 Db 317339 AGTGGCAGCAGTACTGTGTGTCAGTTATA-----AAGGTATTAATCAATTATATGATATG 317392
 Qy 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGluIleAlaHisCysTyrAla 188
 Db 317393 AAGTATATCTAATTTA-----TACATTTTAAACTACGCTGATTCGCAATATGAACTTA 317446
 Qy 189 GluLysValIleGlySerGlyPheAspValAlaIleThrAlaIleTyr---GlyLeuIleVal 207
 Db 317447 CAAGTTTAAAGTTCAATGCGGA---GATATGCTGTGAGTGATATATAGTGGATGATGAGG 317503
 Qy 208 TyrArgArgPheGluProAlaLeuIleAsnAspValPheGluValLeuGluSerAspPro 227
 Db 317504 TATAGTACTTTTATGATCATGATGAGTAAAGCATCAAAATTTGAA----- 317545
 Qy 228 GluLysPheProThrGluLeuLysLysLeuIleGluSerAsnThrGluGluLysHis--- 246
 Db 317546 -----GATACCTACGCTGTAAGAGTTTAACTCAAAACTGGCTGCAATTGCAATC 317596
 Qy 247 GluArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValIleGlyGlySer 266
 Db 317597 GAACCATTAACAAGCACTGAAATAATGAAAGTACTTATGCGTTGAGCTGCTCACCGGCG 317656
 Qy 267 GluThrProLysLeuValSerArgValLeuGluIleTyrLysGluLysProGluGluSer 286
 Db 317657 TCATACACCACTTGTAGTCAAGAGTGAACCTTTGAAA----- 317695
 Qy 287 SerValValTyrAspGluLeuAsnSerAlaAsnLeuGluPheMetLysGluLeuArgGlu 306
 Db 317695 ----- 317695
 Qy 307 MetArgGluLysTyrAspSerAspProGluThrTyr-----IleLysGluLeuAsp 323
 Db 317696 -----TCACATCTTCATTTTACGCGTCTTCAAGATTCACACT 317737
 Qy 324 HisSerValGluProLeuThrValAlaIleLys---AsnIleArgLysGlyLeuGluAla 342
 Db 317738 CATTGCTTTGAAAACCTTATTCATCTTTTAAACAATTAACATTAAGGTGTGCAAAAG 317797
 Qy 343 LeuThrGluLysSerGluValProIleGlu-----ProAspVal 355
 Db 317798 ATGGTCCGCGCAAAATCGTACAAATTATTCACGATGATTAAGAAAGCTACAGTTATATA 317857
 Qy 356 GluThrGlu---LeuLeuAspArgCysGluGluIleProGlyCysValGly 371
 Db 317858 GAAACTGAAAACCTAAATATTTGTGTGATATTCGCGAAAATATACACCGT-----GCA 317911
 Qy 372 GlyValValProGlyValGlyCylTyrAspAlaIleAlaValLeuValLeuGluAsnGln 391
 Db 317912 TCTAAACATCAGCGCGCTGTGTGTGAGACTGT---GGATATCAATTATCAATAAGAT 317968
 Qy 392 ValGlyAsnPheLys 396
 Db 317969 GTAGATTAAGAAAAA 317983
 RESULT 39
 AL591973 200050 bp DNA linear BCT 06-JUN-2002
 DEFINITION AL591973 AL591824 Listeria monocytogenes strain EGD, complete genome, segment 1/12.
 ACCESSION AL591973.1 GI:16409359
 VERSION AL591973.1 GI:16409359
 KEYWORDS
 SOURCE Listeria monocytogenes
 ORGANISM Listeria monocytogenes
 Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 REFERENCE 1 Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
 AUTHORS

FEATURES
 source
 gene
 RBS
 CDS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 FEATURES
 source
 1. 200050
 Location/Qualifiers
 /organism="Listeria monocytogenes"
 /mol_type="genomic DNA"
 /strain="EGD-e"
 /db_xref="taxon:1639"
 305..1673
 /gene="dnaA"
 305..310
 /gene="dnaA"
 318..1673
 /gene="dnaA"
 /codon_start=1
 /transl_table=1
 /product="Chromosomal replication initiation protein DnaA"
 /protein_id="CAC98216.1"
 /db_xref="GI:16409360"
 /db_xref="SPTREMBL:O8YAN2"
 /translation="MOSIEDTWQETLQIVKMSKPSYDTVMKSTTAHSELEGNFTIIS
 APNNFVDMWEKSTYQPIANTLQETRLPDPVREIDSEQENFEYTVKPNPALDEG
 IEIKHMLNPPRYVDFTVIGSGNFPAAASLAVAEAPKAPNPLFIYGVGLGTHLM
 HAVGHVVOOHKDNKAWYLSSEKPTNFISIRNKTEBEPKRYANDVLLIDIOPL
 AGKGTQEEFPHTNTLYDEKQIISDRPKKIPLELRBSRFNGLTIDTPPD
 LERIALIRKAKADGIDIPREVMYLRANQIDNIRLBEGLIRVAVASIPVKNQDTA
 GLAEALDKIIPSSKQVITISGLOAVGEYFHLDPFAKKRTKSIAFPRQIAMLV
 SRELTDASLPRTIGDEFGRDHTVIHAHEKISQLKTDQVLKNDLABIEIKLRAQNM
 F"
 1856..3062
 /gene="dnaN"
 1856..1860
 /gene="dnaN"
 1867..3012
 /gene="dnaN"
 /codon_start=1
 /transl_table=1
 /product="DNA polymerase III, beta chain"
 /protein_id="CAC98217.1"
 /db_xref="GI:16409361"
 /db_xref="SPTREMBL:O8YAN1"
 /translation="MKFVIBRDRLVQAVNVEVTRASRTTPIITGIRIVNDEGVTL
 TGSDDISIRAFIPLINDRIVIVESRGIVIOSKRGDIIVRLPENNVEIETSYN
 QTNISQASPTLNGLDPMRYPKLPETDQKTIKIPINLKNIRQTVFAYSAIEVRP
 VLTGVNMTIKENKLSAVATDSHRLALEIPLETDIDENYINIVIPGKLSLENKLDDA
 SESIENTLANNQILFKKDLIPYSLREGSPYDTSRLIPDTKSELVINSKALQALD
 RASLILARENENNVKIMLTLENGQVEVSNSPEVGVNSGVNSOSFTGESEIKISFNKGY
 MMDLRAFEGBDDQISFSGTMRPVLAKPDAANPNEILQILTPRTY"
 3030..3062
 /gene="dnaN"
 3104..4464
 /gene="lmo0003"
 3104..3109
 terminator
 gene
 RBS

Baquero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T.,
 Charbit, A., Chetouani, F., Couve, E., de Darvar, A., Dehoux, P.,
 Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L.,
 Dueset, O., Ertan, K. D., Fejhi, H., Portillo, F. G., Garrido, P.,
 Gautier, L., Goebel, M., Gomez-Lopez, N., Hain, T., Hauf, J.,
 Jackson, D., Jones, L. M., Kaerst, U., Kretz, J., Kuhn, M., Kunst, F.,
 Kurapkat, G., Madueno, E., Maitouram, A., Vicente, J. M., Ng, E.,
 Nedjari, H., Nordiek, G., Novella, S., de Pablo, B., Perez-Diaz, J. C.,
 Purcell, R., Remmel, B., Rose, M., Schleuter, T., Simes, N.,
 Trierer, A., Vazquez-Bolaj, J. A., Voss, H., Wehland, J. and Cossart, P.
 Comparative genomics of *Listeria species*
 Science 294 (5543), 849-852 (2001)

2 (bases 1 to 200050)
 11679669
 Glaeser, P., Frangeul, L. and Rusniok, C.
 Direct Submission
 Submitted (06-JUN-2001) Glaeser P., Institut Pasteur, Genomique des
 Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris
 Cedex 15, FRANCE
 E-mail: pglaser@pasteur.fr
 Phone: +33 1 45 68 89 96, Fax: +33 (0) 1 45 68 87 86.

Db 15020 TGGACCTTTACACCGGACCAATTATATCGGACACTTCTCTGAAA--TCAGAGAGA 15076
 Qy PheAspLeu-----GlutllelleIetyrSerAspProGlyTyrrHisserGlnu 118
 Db 15077 ATCGAGCTTACACCTGTGAAATGATATAT----- 15106
 Qy AspThrGluThrLyethSerSerhenglyGlyuLythrPheLeuTyrrHisserArgala 138
 Db 15107 GAAACAGAAATTATATCGACCAATCTGTGCA----- 15136
 Qy IleThrGluValGlyuLyethSerSerArgalaGlyLeuValSerValala 158
 Db 15137 -----AAATATGACTAGGTTCAGGCGCGGACCGGCGGCTTCTGTAATC 15181
 Qy ThrSerLeuSerHisPheIleProAsnValIleSerThrAsnLyAspIleLeuHis 178
 Db 15182 AACGGCTATATGACGAATTTATTCAGAAATATCTATCTTAAATA-----TTC 15232
 Qy AsnValAlaGlnIleAlaHisCysTyrrAlaGlnLyLeuGlySerGlyPheAspVal 198
 Db 15233 AAACAGCTGACACTTCCATTTAGCTTCAAGGAAAC--GTTTCATGTGTGATATTT 15289
 Qy AlaThrAlaIleTyrr-----GlyLeuIleValTyrrArgArgPheGlnProAlaLeuIleAsn 217
 Db 15290 GCATCTGTATGTACGCGTGTGATGCGATGCGATACGACCTTGTATCAAGATGAGATGAG 15349
 Qy AspValPheGlnValLeuGluSerAspProGluLyPheProThrGluLeuLySerLeu 237
 Db 15350 -----CATCGTTTACCTTAAATCGCTCGAATGTGTTATGAAAGAGCCGCGCAATG 15403
 Qy IleGluSerAsnTrpGluGlyuLyshiGlyuArgCysThrLeuProTyrrGlyIleLySerLeu 257
 Db 15404 CTTCGAATC-----GAAACATTAGAGAAACAGACGCGACCTTTCTGTGTA----- 15448
 Qy LeuMetGlyAspValLyGlySerGluThrPro-----LysLeuValSer 273
 Db 15449 -----GGTGGACGGGACACCTGTAAGTACCGGAGAAATAGTATTCG 15490
 Qy ArgValLeuGlnTrpLySerGlyuLyPProGluGluSerSerValValTyrrAspGlnLeu 293
 Db 15491 CAATTCATCGCTTTTAA-----CAAGAGATAGCAAGAAATTAACCAACATTTT 15538
 Qy AsnSerAlaAsnLeuGlnPheMetLySerGluLeuArgGluMetArgGlyuLyTyrrAspSer 313
 Db 15539 TTAACAGAAATTAATGAGATTAATGAGCAATTAATTAATCAAGCTTCCATACGAAAGACGAG 15598
 Qy AspProGluThrTyrrIleLySerGluLeuAspHisSerValGluProLeuThrValAlaIle 333
 Db 15599 -----GAATTCCTCTATTCACCTATTT 15619
 Qy LysAsnIleArgLyGlyLeuGlnAlaLeuThrGlnLySerGluValProIleGluPro 353
 Db 15620 AAAGAAATCGTCGATCTTCAAGAACTTGGAAACAAAGCTGGCGGTAATTAATTAACA 15679
 Qy AspValGlnThrGlnLeuLeuAspArgCysGlnGlnIleProGlyCysValGlyIleVal 373
 Db 15680 AGCTTGCTAAAGAAAGACTCGCGGACCTCAGCTGAAACATGCGGCGC-----GAGGAGAAA 15733
 Qy ValPProGlyAlaGlyGlyTyrrAsp 381
 Db 15734 TCATCTGTGCTGTGCGCGGAGAC 15757
 RESULT 40
 AX638811 349980 bp DNA linear PAT 21-FEB-2003
 LOCUS AX638811
 DEFINITION Sequence 1 from Patent WO0101118.
 ACCESSION AX638811
 VERSION AX638811.1 GI:28474425
 KEYWORDS
 SOURCE Listeria monocytogenes
 ORGANISM Listeria monocytogenes
 Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 REFERENCE 1

AUTHORS Buchrieser, C., Frangeul, L., Couve, E., Rusniok, C., Fsihi, H.,
 Denoux, P., Dussurget, O., Chetouani, F., Nedjari, H., Glaser, P.,
 Kunst, F., Cossart, P., Daniels, W., Kneif, J., Kuhn, M.,
 Ng, E., Vazquez-Boland, J., Dominguez-Bernal, G., Garrido-Garcia, P.,
 Trierer-Martinez, A., Amend, A., Chakraborty, T., Domann, E., Hain, T.,
 Berche, P., Charbit, A., Durant, L., Perez-Diaz, J. C., Baquero, F.,
 Garcia del Portillo, F., Gomez-Lopez, N., Madueno, E., de Pablos, B.,
 Wehlund, J., Kaerst, U., Entian, K. D., Hauf, J., Rose, M., and Voss, H.
 Listeria monocytogenes genome, polypeptides and uses
 Patent: WO 0101118-A 1 11-APR-2001;
 JOURNAL PASTEUR (FR)
 INSTITUT PASTEUR (FR)
 FEATURES
 source location/Qualifiers
 1. 349980
 /organism="Listeria monocytogenes"
 /mol type="Genomic DNA"
 /db xref="taxon:1639"
 /note="seq 1 for big; original length: 2944528-splittd
 into 10 <223> seq 0001 : From 0.000.001 to 0.349.980
 length: 349.980 <223>seq 2855 : From 0.300.001 to
 0.649.980 length: 349.980 <223>seq 2856 : From 0.600.001
 to 0.949.980 length: 349.980 <223>seq 2857 : From
 0.900.001 to 1.249.980 length: 349.980 <223>seq 2858 :
 From 1.200.001 to 1.549.980 length: 349.980 <223>seq 2859
 : From 1.500.001 to 1.849.980 length: 349.980 <223>seq
 2860 : From 1.800.001 to 2.149.980 length: 349.980
 <223>seq 2861 : From 2.100.001 to 2.449.980
 length: 349.980 <223>seq 2862 : From 2.400.001 to 2.749.980
 length: 349.980 <223>seq 2863 : From 2.700.001 to
 2.944.528 length: 244.528"
 BASE COUNT 115218 a 58816 c 74738 g 101208 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.0387 length: 349980
 Score: 170.50 Matches: 96
 Percent Similarity: 38.92% Conservative: 55
 Best Local Similarity: 24.74% Mismatches: 144
 Query Match: 7.65% Indels: 93
 DB: 6 Gaps: 19
 US-10-069-062-7 (1-432) x AX638811 (1-349980)
 Qy 8 ProGlyuysAlaPheLeuAlaGlyGlyTyrrLeuValLeuGluProIleTyrrAspAlaTyrr 27
 Db 14831 CCGGAAATTAATATGTTGCTGCGATATATGACATTTAGAGATCAGGTCAATACGGCTATT 14890
 Qy 28 ValThrAlaLeuSerSerArgMetHisAlaValIleThrProLyGlyThrSerLeuLys 47
 Db 14891 CTAAACCGCG-----GTTAACCGGTATATTAACG-----CTAACTTAGAG 14929
 Qy 48 GluSerArgIleLyIleSerSerProGlnPheAlaAsnGlyGluTrpGluTyrrHisIle 67
 Db 14930 GATAGTAA-----CGAATGAATTAATGATTCACATTAAT 14965
 Qy 68 SerSerAsnThrGluLyPProArgGluValGlnSerArgIleAsnPro----- 83
 Db 14966 GAAATTCAGTTTATATGCGCA-----CTTGGCGGCGCACTTAACCTGACGAGAAAT 15019
 Qy 84 -----PheLeuGluAlaThrIlePheIleValLeuAlaTyrrIleGlnProThrGlnAla 101
 Db 15020 TGGACCTTTACAGCGGAAAGCAATTAATATCGGCAACTTCTCGTAAA--TCAGAGAGA 15076
 Qy 102 PheAspLeu-----GlutllelleIetyrSerAspProGlyTyrrHisserGlnu 118
 Db 15077 ATCGAGCTTACACCTGTGAAATGATATAT----- 15106
 Qy AspThrGluThrLyethSerSerhenglyGlyuLythrPheLeuTyrrHisserArgala 138
 Db 15107 GAAACAGAAATTATATCGACCAATCTGTGCA----- 15136
 Qy IleThrGluValGlyuLyethSerSerArgalaGlyLeuValSerValala 158
 Db 15137 -----AAATATGACTAGGTTCAGGCGCGGACCGGCGGCTTCTGTAATC 15181

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: August 15, 2003, 12:44:33 ; Search time 82 Seconds
(without alignments)
2325.336 Million cell updates/sec

Title: US-10-069-062-7
Perfect score: 2230
Sequence: 1 MSKASAPGKAFIAGYLVL.....DLEQTEGVLEKPHDYIGL 432

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 569978 segs, 220691566 residues
Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n_model -DEV=x1h
-O=/cgn2_1/USPTO_spool/US10069062/runat_07082003_161503_4139/app_query.fastea_1.583
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rm1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=pico -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10069062@cgn1.1.56 @runat_07082003_161503_4139 -NCPUS=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Issued Patents NA:*
- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
 - 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
 - 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
 - 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
 - 5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:*
 - 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	181.5	8.1	1089	US-09-134-001C-2518	Sequence 2518, Ap
2	177.5	8.0	1098	US-09-107-532A-2936	Sequence 2936, Ap
3	139.5	6.3	312	US-09-313-294A-5178	Sequence 5178, Ap
4	117	5.2	11887	US-08-961-527-146	Sequence 146, App
5	113	5.1	1847	US-08-858-207A-64	Sequence 64, App1
6	111.5	5.0	2223	US-08-257-073-4	Sequence 4, App1
7	108.5	4.9	8257	US-09-595-684B-30	Sequence 30, App1
8	108.5	4.9	8503	US-09-620-312D-130	Sequence 130, App
9	108	4.8	1230025	US-09-198-452A-1	Sequence 1, App1
10	105	4.7	2451	US-09-107-532A-208	Sequence 208, App
11	105	4.7	7163	US-08-961-527-67	Sequence 67, App1
12	101	4.5	13977	US-09-484-970B-60	Sequence 60, App1

13	101	4.5	37948	US-09-251-645-11	Sequence 11, App1
14	99	4.4	580073	US-08-545-528D-1	Sequence 1, App1
15	98	4.4	536165	US-09-214-808-1	Sequence 1, App1
16	97.5	4.4	1521	US-09-134-001C-1546	Sequence 1546, Ap
17	96.5	4.3	1044	US-09-495-406-12	Sequence 12, App1
18	96.5	4.3	2793	US-08-347-563A-1	Sequence 1, App1
19	96.5	4.3	2793	US-08-485-942A-1	Sequence 1, App1
20	96.5	4.3	2793	US-08-488-214A-1	Sequence 1, App1
21	96.5	4.3	2793	US-08-488-208A-1	Sequence 1, App1
22	96.5	4.3	2793	US-08-483-211A-1	Sequence 1, App1
23	96.5	4.3	2793	US-08-488-223A-1	Sequence 1, App1
24	96.5	4.3	2793	US-08-438-431A-1	Sequence 1, App1
25	96.5	4.3	2793	US-08-488-225A-1	Sequence 1, App1
26	96.5	4.3	4295	US-09-620-312D-338	Sequence 338, App
27	96.5	4.3	11474	US-09-495-406-1	Sequence 1, App1
28	96	4.3	1835	US-09-620-312D-200	Sequence 200, App
29	96	4.3	16995	US-08-961-527-82	Sequence 82, App1
30	95.5	4.3	1029	US-09-107-532A-2010	Sequence 2010, Ap
31	95.5	4.3	2608	US-09-154-750A-75	Sequence 75, App1
32	95	4.3	11887	US-08-961-527-146	Sequence 146, App
33	94.5	4.2	1710	US-08-969-046-3	Sequence 3, App1
34	94.5	4.2	2840	PCT-US93-06251-64	Sequence 64, App1
35	94	4.2	1941	US-09-107-532A-2059	Sequence 2059, Ap
36	94	4.2	3023	US-09-308-022-4	Sequence 4, App1
37	94	4.2	11832	US-08-416-603-1	Sequence 1, App1
38	94	4.2	580073	US-08-545-528D-1	Sequence 1, App1
39	93.5	4.2	843	US-09-328-352-1259	Sequence 1259, Ap
40	93.5	4.2	5517	US-09-120-663-1	Sequence 1, App1
41	93	4.2	7393	US-09-620-312D-372	Sequence 372, App
42	92.5	4.1	1926	US-09-590-020-1	Sequence 1, App1
43	92.5	4.1	4868	US-08-139-937-12	Sequence 12, App1
44	92.5	4.1	4868	PCT-US93-11310-12	Sequence 12, App1
45	92.5	4.1	6773	US-09-166-350-27	Sequence 27, App1

ALIGNMENTS

RESULT 1
US-09-134-001C-2518
; Sequence 2518, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2518
; LENGTH: 1089
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2518

Alignment Scores:

Pred. No.: 6,48e-12 Length: 1089
Score: 181.50 Matches: 99
Percent Similarity: 36.74% Conservative: 52
Best Local Similarity: 24.09% Mismatches: 131
Query Match: 8.14% Indels: 129
DB: 4 Gaps: 18

US-10-069-062-7 (1-432) x US-09-134-001C-2518 (1-1089)

CY 7 AAlProGlyLysAlaPheLeuAlaGlyGlyTyrLeuValLeuGluProlIleTyrAspAla 26
DB 28 GCCCCCGGAAACCTTATATTCGACGAGATGTCAGTAACCGAACCAGATATATATCT 87

27 TyrValThrAlaLeuSerSerIleArgMetHisAlaValIle----- 39
Db 88 ATTCTATTGCGAGTAATGCGCTTTGTAACGGCACATTAAGCGCTCAATAAGTTGAA 147
QY 40 -----ThProLysGlyThrSerLeuLysGlySer 49
Db 148 GGTAGATTCAATCCAAAACATTACTATTGAAACCAAGTTAAATTGACCGTATGAAGAT 207
QY 50 ArgIleLysIleSerSerProGlnPheAlaGnglyGluTrpGluTyrHisIleSerSer 69
Db 208 AGAATTGAATCTCAAGATGTTCAAGCTGCTTAAGCA----- 243
QY 70 AsnThrGluLysProArgGluValGlnSerArgIleAsnProPheLeuGluValAlaThrIle 89
Db 244 -----CGAATAATGTTGGT 258
QY 90 PheIleValLeuAlaTyrIleGlnProThrGluAlaPheAspLeuGluIleIleTyr 109
Db 259 AAGCATATGAAAGTGTGTAACAGTATGTCGCGAGTTGCATATGAATTTA----- 309
QY 110 SerAspProGlyTyrHisSerGlnLysAspThrGluThrLysThrSerSerAsnGlyGlu 129
Db 310 -----AAGCACTTTCATTTAAACCATTTGATAGTACCTTGAACAATTACTCTGCT----- 357
QY 130 LysThrPheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySer 149
Db 358 -----CAGAAAGTACGATTAAGTTCA 378
QY 150 SerAlaGly-----LeuValSerValValAlaThrSerLeuLeuSerHisPheIleProAsn 168
Db 379 AAGCGCGCGCTTTTACTATCTGTGTT-----AAGCTTTGAATGATTTCTAATGCT--- 429
QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
Db 430 ---TTGGAATTTATCAAACTTATATTATTTAAATAGCTGTAATTCGAATATGAATTA 486
QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyr-----GlyLeuIleVal 207
Db 487 CAAGCTTAAAGTTCATGTGCT---GATATTCGCGTTAGTGTCTACAGTGGTGGCTTGA 543
QY 208 TyrArgArgPheGlnProAlaLeuIleAsnAsp---ValPheGlnValLeuGlnSerAsp 226
Db 544 TATAGTACGTTGAC-----CATGATCGGTGAACAACGCAAAAGCAAGAA--- 588
QY 227 ProGlnLysPheProThrGluLeuLysLysLeuIleGlnSerAsnTrpGlnLysHis 246
Db 589 -----ACATCGGTGAATGATGTTTGGAAAAAATATGGCCAGCGCTTACAT 633
QY 247 ---GluArgCysThrLeuProTyrGlyIleLysLeuLeuMetClyLysAspValLysGlyGly 255
Db 634 ATCGAACCTTTACAGACCTCCCAAAATATGGAAGCTCTTATGGATGAGCATGGGTCCCA 693
QY 266 SerGlnThrProLysLeuValSerArgValLeuGlnTrpLysLysGluLysProGlnGlu 285
Db 694 GCTTCTTCCTCCACACTTAAGTGAAGTGAAGTCAAACTTTAAA----- 735
QY 286 SerSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGluLeuArg 305
Db 735 ----- 735
QY 306 GluMetArgGlnLysTyrAspSerAspProGlnIleThrTyrIleLysGluLeuAspHisSer 325
Db 736 -----TCAGATCCCAAGATTTTATAGTGATTTTATAGATCAATCT 774
QY 326 ---ValGluProLeuThrValAlaIleLys-----AsnIle----- 336
Db 775 CATGCTTGCTGTAAGAAAGTTAAACCAAGCTTTTAAACATAATATCAAAAGTGTCCA 834
QY 337 ---ArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValPro 350
Db 835 AAGATGATACGTATAAACAAGACGTATATTATTCATCTATGATGAATACGAACATCAGTTGAA 894
QY 351 IleGluProAspValGlnThrGlnLeuLeuAspArgCysGlnIleLysProGlyCysVal 370

```

Db      895 ATTGAAACAGTAAAGCTTAAAGAAATTTATGTAATTCGCTGATAAAGACGGTGGC----- 948
Qy      371 G1yG1yValValProG1yAlaG1yG1yTYrLeuValLeuG1yProIleTYrAsp 381
Db      949 GCTTCTAAACTTCAGGTGCTGGTGGTGGCGCAT 981

RESULT 2
US-09-107-532A-2936
/ Sequence 2936, Application US/09107532A
/ Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
FAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 2936:
SEQUENCE CHARACTERISTICS:
LENGTH: 1098 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...1098
SEQUENCE DESCRIPTION: SEQ ID NO: 2936:
US-09-107-532A-2936

Alignment Scores:
Pred. No.: 1.98e-11 length: 1098
Score: 177.50 Matches: 97
Percent Similarity: 39.55% Conservative: 62
Best Local Similarity: 24.13% Mismatches: 144
Query Match: 7.96% Indels: 99
DB: 4 Gaps: 18

US-10-069-062-7 (1-432) x US-09-107-532A-2936 (1-1098)
Qy      6 serialProG1yValAlaPheLeuAlaG1yG1yTYrLeuValLeuG1yProIleTYrAsp 25
Db      25 TCTGACACAGGCAACTCTATATTGCGCGAATATGCAAGTTGTGAAACAGGCGCATCA 84

```

```

QY 26 AlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLysGlyThrSer 45
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85 GCAGTATATGCGTCACTGATCAATTCGGACAGTCACGTGA----- 126

QY 46 LeuLysGlnSerArgIleLysIleSerSerProGlnPheAla----- 59
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 127 -----GAATTCGACGAAAGATCGGAAGATCCAAATCTGCACAAATATATGCGGATGCGCT 180

QY 60 -----AsnGlyIuTrpGluTyrHisIleSerSerAsnThrGluLys 73
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 GTAAGCTTGACAAAGACGCAAGCAAGAAATTGATTGATTA----- 222

QY 74 ProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeu 93
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 -----CGGGAATATCCCTTT-----CATTAATATCCCTT 249

QY 94 AlaTyrIleGlnProThrGluAlaPheAspLeuGlu-----IleIleIleTyrSerAsp 111
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 250 GGTGCTATTCGCTTGACATAAGATGACAAAGAAAGATGACAAAGAAATCTCTTTA----- 300

QY 112 ProGlyTyrHisSerGlnLysAspThrGluThrIleThrSerSerAsnGlyGluLysThr 131
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 301 -----TCCTTTATGATCTGAAGTAAACAGTGAAGTGAACAGTTCAAATGCG----- 348

QY 132 PheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAla 151
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 349 -----CGGAATATGATGATGGGTTCTAC----- 372

QY 152 GlyLeuValSerValAlaThrSerLeuSerHisPheIleProAsnValIleSer 171
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 GGTGCGGTCGCTGCAACGCTGCAAGCGCTGAAGTGTGTTTAC-----CGCTTGAAAT 426

QY 172 ThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysValAlaGlnLysLys 191
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 427 TTATCTCAGCTTGAGATTTTCAAGATGACGACATGACCAATTTAGACAGTTAAGATTAAT 486

QY 192 IleGlySerGlyPheAspValAlaThrAlaIleTyr-----GlyLeuIleValTyrArgArg 210
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 487 ---GGTTCGCGGAGACATGCGCTGCTGATGCTGTATGCGGTGATGCGCTTCTTCACAC 543

QY 211 PheGlnProAlaLeuIleAsnAspValPheGlnValLeuGlnSerAspProGluLysPhe 230
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 544 TTCGACCAATCTTGGCTCCAGAA-----CAAGAAACT 576

QY 231 ProThrGlnLeuLysLeuIleGlnSerAsnTrpGlnGlu-----LysHisGlnLysCys 249
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 577 CAGCAATTCATCAGATGATTAAGTCTGCTGATGCGCAGGTCTATCCATGAGCCATG 636

QY 250 ThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValIleGlyGlySerGlnThrPro 269
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 637 ATTGCTCCGAGATTTACGTTTATGATGCTTGAGACGGTACGCCCTGCTCTACTCT 696

QY 270 LysLeuValSerArgValLeuGlnTrpLysGlyLysProGlnGlnSerSerValVal 289
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 697 GATTGGTCGATCAAGTT-----CACCGTTGAGAGAGATTAATAATGGTG 741

QY 290 TyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGlu-----LeuArg 305
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 742 GCTTATCAGCTT-----TTCTTAAATAATATGACAGATGTGTCAAC 783

QY 306 GluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGlnLeuAspHisSer 325
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 784 GAAATGATCAAGGCTTTAAAGAAATATATGTAAGCTTGATTCACACAGATG----- 834

QY 326 ValGluProLeuThrValAlaIleLysAsnIleArgLysGlyLeuGlnAlaLeuThrGln 345
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 835 -----ATTGCAAAAAACCGACCAATTAAGTCAATTTATCTGCA 873

QY 346 LysSerGlnValProIleGlnProAspValGlnThrGlnLeuAspArgCysGlnGlu 365
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 874 ATCACTGGGGTGTGATCGAAAGCGCTGCTTGAACAAATGTGTATTAATTTAGCTAAACAG 933

```

```

QY 366 IleProGlyCysValGlyValValProGlyAlaGlyGlyTyrAspAlaIleAlaVal 385
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 934 TATGAGAGAGCGCAAAATCT-----TCTGCTGACAGGAGGGGCGATTCGGAAATCGTA 987

QY 386 LeuVal 387
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 988 ATGTGT 993

RESULT 3
US-09-313-294A-5178
; Sequence 5178, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 5178
; LENGTH: 312
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6476212 700349731H1
; NAME/KEY: unsure
; LOCATION: 241, 245, 249
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-5178

Alignment Scores:
Pred. No.: 8.85e-08 Length: 312
Score: 139.50 Matches: 34
Percent Similarity: 51.89% Conservative: 21
Best Local Similarity: 32.08% Mismatches: 40
Query Match: 6.26% Indels: 11
DB: 4 Gaps: 3

US-10-069-062-7 (1-432) x US-09-313-294A-5178 (1-312)

QY 192 IleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyrArgArgPhe 211
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 ATTGCAAGTGTGTTGATGTTAGTGCCTGCTGTATGGAAGTCAACGCTATGTAAAGTTT 62

QY 212 GlnProAlaLeuIleAsnAspValPheGlnValLeuGlnSerAspProGluLys----- 229
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 TCTCCA-----GAATTAATCTGCTCTCAAGCTTACAGCTGCGGACT 101

QY 230 ---PheProThrGlnLeuLysLeuIleGlnSerAsnTrpGlnGluLysHisGlnLys 248
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 102 TTCCTCCAGATGATGATCAAGATTTCTTACAAAGAGTGGATCATGAAATTAACAG 161

QY 249 CysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspVal-----LysGlyGlySerGlu 267
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 162 TTCATTAATCTCTCTGATGACCTTCTCTTGGGAAACCTGAAACCGAGAGATCATCT 221

QY 268 ThrProLysLeuValSerArgValLeuGlnTrpLysGlyLysProGlnGlnSerSer 287
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 ACTTCATCAATGCTTGATGATNTGTAAACGTCGGCAGAAAGTGTGACCTCGAATAATGCAA 281

QY 288 ValValTyrAspGlnLeu 293
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 282 GATACATGAGATAAAGCTG 299

RESULT 4
US-08-961-527-146/c
; Sequence 146, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:

```

APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 146:
SEQUENCE CHARACTERISTICS:
LENGTH: 11887 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-146
Alignment Scores:
Pred. No.: 0.0174 Length: 11887
Score: 117.00 Matches: 95
Percent Similarity: 34.91% Conservative: 53
Best Local Similarity: 22.41% Mismatches: 146
Query Match: 5.25% Indels: 130
DB: 4 Gaps: 20
US-10-069-062-7 (1-432) x US-08-961-527-146 (1-11887)
QY 9 GtlyleAlaPheleuAlaGlyGlyTyrLeuValLeuGluProIleTyrAspAlaTyrVal 28
DB 5640 GGAAGAACTCTATGCGCAGGTGAATGCTATTGAGCCAGCGAGTTAGCTTGATA 5581
QY 29 ThrAlaLeuSerSerArgMetHisAlaValIleThrProIleGlyHisSerLeuGlu 48
DB 5580 AAGGATATTCCTCCATCTATATGAGCGCTGAGATT-----GCTTTTCTGAC 5536
QY 49 Ser---ArgIleIleSerSerProGlnPheAlaAsnGlyLutrpGluTyrHisIle 67
DB 5535 AGCTACCGATCTATTCAGATATGTTGATTGCA-----GCTTTTCTGAC 5500
QY 68 SerSerAsnThrGluLysProArgGluValGlnSerArgIleAsnPro-----Phe 84
DB 5499 -----GTGGACTTAAGCGCTATCTGACTGACTATGCTTG 5467
QY 85 LeuGluAlaThrIlePheIleValLeuAlaTyrIle-----GlnProThrGlu 100
DB 5466 ATTCAGAAACGATTCCTTGATGGAGACTTCCTGCTGCTTCGTGCTGAAATTTAA 5407
QY 101 AlaPheAspLeuGluIleIleIleTyrSerAspProGlyTyrHisSerGlnAspThr 120
DB 5406 CCTTTTCTGTAAGATCTCT-----5386
QY 121 GluThrLysThrSerSerAsnGlyGluLysThrPheLeuTyrHisSerArgAlaIleThr 140
DB 5386

DB 5385 -----GCCAAAAATGGAACGA 5371
QY 141 GluValGluLysThrGlyLeuGlySerSerAlaGlyLeuValSerValAlaThrSer 160
DB 5370 GAAGGAAATCTTGGCTGTAGGCTGAGCGAGCGGCTGCTGCTGCTGCTGCTGCTG 5311
QY 161 LeuLeuSerHisPheIleProAsnValIleSerThrAsnLysAspIleLeuHisAsnVal 180
DB 5310 TTACTGGCTCTGTAT-----GATGTTCTGTGATCAGAGCTCTGTTCAAGCTG 5260
QY 181 AlaGlnIleAlaHisCysTyrAlaGlnLysIleGlySerGlyPheAspValAlaThr 200
DB 5259 ACTAGC---GCTGTCTGTCTTAAGCAGAGAGCAATGTTCCATGCGCAGCTTGTCT 5203
QY 201 AlaIle---TyrGlyLeuIleValTyrArgArgPheGlnProAlaLeuIleAsnAspVal 219
DB 5202 ATTGGGACAGAGATTGGTTGTTCTTACCGCTATTTGATCGC-----CAGAGGTG 5152
QY 220 PheGlnValLeuGluSerAspProGluLysPheProThrGluLeuLysLysLeuIleGlu 239
DB 5151 GCTGCTGCTTGAAGAGAGA-----AACTGGCCAGATTCTCGAG 5110
QY 240 SerAsnTrpGluGluLysHisGluArgCysThrLeuProTyrGlyIleLysLeuLeuMet 259
DB 5109 CCGTATTTG-----GGCTTTTCAATT-----5089
QY 260 GlyAspValLysGlyLysSerGluThrProLysLeuValSerArgValLeuGlnTrpLys 279
DB 5088 TCACAGTGAACCACTTATGATGTATTTCTTATGAGCA-----TGACCC 5041
QY 280 LysGluLysProGluGluSerSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGln 299
DB 5040 AAGGAAGTGGCTGTATGAGTCAGTCATGTCGACCAATCAACAAATATCATCAAAAT 4981
QY 300 PheMetLysGluLeuArgGluMetArgGluLysTyrAspSerAspProGluThrTyrIle 319
DB 4980 TTT-----4978
QY 320 LysGluLeuAspHisSerValGluProLeuThrValAlaIleLysAsnIleArgLysGly 339
DB 4977 -----TTACTTCTCAAAAGAACGGTACTTCTCTGTTGAAAGCTTGAAACGGGG 4924
QY 340 ---LeuGlnAlaLeuThrGlnLysSerGluValPro-----IleGlu 352
DB 4923 AATCAGAAAGATTTATGATCATGATGAGTACAGCAGCAGCTTTGAGAGCTTGAGT 4864
QY 353 ProAspValGlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyGly 372
DB 4863 ACAGATATTACACCGCTTGTCTTACAGCTTGAAGAAAGCCAGTCAGATTGTCAGACC 4804
QY 373 ValVal-----ProGlyValaGlyLysTyrAspAlaIleAlaValLeuValLeuGluAsn 390
DB 4803 GTTGCCAGAGTAGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4744
QY 391 GlnValGlyAsnPheLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
DB 4743 CAA-----TCAACCAAAACCTTAATAAAAT-----CGT 4717
QY 411 TrpValAspLeu 414
DB 4716 TGGGCGGATCTG 4705
RESULT 5
US-08-858-207A-64
Sequence 64, Application US/08858207A
Patent No. 6348328
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: No. 6348328e1 Compounds

NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 1847 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-858-207A-64
Alignment Scores:
Pred. No.: 0.00245 Length: 1847
Score: 113.00 Matches: 93
Percent Similarity: 35.27% Conservative: 59
Best Local Similarity: 21.58% Mismatches: 135
Query Match: 5.07% Indels: 144
Gaps: 20
US-10-069-062-7 (1-432) x US-08-858-207A-64 (1-1847)
QY 9 GlyValAlaPheLeuAlaGlyGlyTyrLeuValLeuGluProIleTyrAspAlaTyrVal 28
DB 903 GGAAGAACTCTATGCGAGGTAATATGCTATTATTTAGAGCCAGGCGACGTTAGCTTGATA 962
QY 29 ThrAlaLeuSerSerArgMetHisAlaValIleThrProIlyGlyThrSerLeuIlyGlu 48
DB 963 AAGGATATTCCTCATATATGAGGCTGAGATT-----GCTTTTCTTAC 1007
QY 49 Ser---ArgIleTyrSerSerProGlnPheAlaGlnGlyGluTyrGluTyrHisIle 67
DB 1008 AGCTACCGATCTATTCAGATATGTTGATTTCGA----- 1043
QY 68 SerSerAenThrGluLysProArgGluValGlnSerArgIleAenPro-----Phe 84
DB 1044 -----GTGACCTTAAGGCCCAATCCGACTACAGCTTG 1076
QY 85 LeuGluAlaThrIlePheIleValLeuAlaTyrIle-----GlnProThrGlu 100
DB 1077 ATTCAGAAAGCATGCTTTGATGGAGACATTCCTCGCTTCGCGGTACAGATTTAAGA 1136
QY 101 AlaPheAspLeuGluIleIleIleTyrSerAspProGlyTyrHisSerGlnGluAspThr 120
DB 1137 CCTTTTCCCTAAATAATCTGT----- 1157
QY 121 GluThrIlyThrSerSerAsnGlyGluLysThrPheLeuTyrHisSerArgAlaIleThr 140
DB 1158 -----GGCAAAATGGAACGA 1172

QY 141 GluValGluTyrThrGlyLeuGlySerSerAlaGlyLeuValSerValAlaThrSer 160
DB 1173 GAGGAGAAAGAACTTGCTAGGCTTCTAGTGCACGCGCTGTTGTTGTCAAGCTG 1232
QY 161 LeuLeuSerHisPheIleProAsnValIleSerThrAsnLysAspIleLeuHisAsnVal 180
DB 1233 TTACTGGCTCTCTAT-----AATCTTCGGTGTATCAGAAATCTGTTTCAAGCTG 1283
QY 181 AlaGlnIleAlaHisCysTyrAlaGlnLysIleGlySerGlyPheAspValAlaThr 200
DB 1284 ACTAC--GCTGCTTCTCAAGCAGAGACAATGGTTCATCGGCGACTTCCTGT 1340
QY 201 AlaIle---TyrGlyLeuIleValTyrArgArgPhe-----GlnProAlaLeuIle--- 216
DB 1341 ATTGGGAGAGAGATTGGTTCCTACCAATGATTTATCCGCAAGCGGCTCTGG 1400
QY 217 -----AenAspValPheGlnValLeuGlnLysAspProGluLysPheProThrGlu 233
DB 1401 TTAGAAAGAAAGAACTTGCGACAGCTTCGAGCGGTATGGGATTTTATATCACA 1460
QY 234 LeuLysLysLeuIleGlnSerAenThrGluGluLysHisGluArgCysThrLeuProTyr 253
DB 1461 GTGAACCACTTAAGATGATTTTC----- 1487
QY 254 GlyIleLysLeuLeuMetGlyAspValLysGlyGlySerGluThrProLysLeuValSer 273
DB 1488 -----TTAGTGGGA----- 1496
QY 274 ArgValLeuGlnTyrLysGlyLysPheGluLysPheGluLysSerValTyrAspGlnLeu 293
DB 1497 -----TGACCAAGAAAGTGGCTGTATCAATCATCATGCGCCACCAATCAAG 1544
QY 294 AenSerAlaAenLeuGlnPheMetLysGluLeuArgIleMetArgLysTyrAspSer 313
DB 1545 CAAATATCATCAATCAAAATTTTAA-----AGTTCC 1574
QY 314 AspProGluThrTyrIleLys-----GluLeuAspHisSer 325
DB 1575 TCAGAAAGAAAGCGGCTTCTTGTGTCAGCTTGACGCGGAGAAAGCCGAAAGATT 1634
QY 326 ValGluProLeuThrValAlaIleLysAsnIleArgLysGlyLeuGlnAlaLeuThrGln 345
DB 1635 ATCGAGCAAGTAAAGTAAAGTACAGCAAG--CTTTTAAAGAGCTTGAGTACA----- 1682
QY 346 LysSerGluValProIleGluProAspValGlnThrGlnLeuLeuAspArgCysGlnGlu 365
DB 1683 -----GATATTTACACGCGCTTGTAGACAGTTGAAGAA 1718
QY 366 IleProGlyCysValGlyValVal-----ProGlyAlaGlyTyrAspAlaIle 383
DB 1719 GCGAGTCAAGATTGCGAGCGCTTGCCAGAGTGTGTGGTGGTGGTACCTGCGC 1778
QY 384 AlaValLeuValLeuGlnAsnGlnValGlyAsnPheLysGlnLysThrLeuGlnAsnPro 403
DB 1779 ATCGCCCGAGATTGTGATGCGCA-----TCAACGAAACCTTAAAAAT--- 1823
QY 404 AspTyrPheHisAsnValTyrTrpValAspLeu 414
DB 1824 -----CGTTGGCGCGATCTG 1838
RESULT 6
US-08-257-073-4
Sequence 4, Application US/08257073
Patent No. 5766597
GENERAL INFORMATION:
APPLICANT: Paoleletti, Enzo
APPLICANT: de Taisne, Charles
APPLICANT: Tine, John A.
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P. C.

STREET: 530 Fifth Avenue, 25th Floor
CITY: New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTMS
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-4
Alignment Scores:
Pred. No.: 0.00502 Length: 2223
Score: 111.50 Matches: 83
Percent Similarity: 35.64% Conservative: 56
Best Local Similarity: 21.28% Mismatches: 110
Query Match: 5.00% Indels: 141
DB: 1 Gaps: 19
US-10-069-062-7 (1-432) x US-08-257-073-4 (1-2223)
QY 126 SerAsnGlyGluuysThrPheLeuYrHisSerArgAlaIleThrGlu----- 141
DB 964 AGTAAGATTATTAATTAATTTCTTAATCAAGTAAGAAATAGAGAAATTTATTAAAC 1023
QY 142 ValGluuysThrGlyLeuGlySerSeraIaGlyLeu-----Val 154
DB 1024 ATAAACAAGATGAATCGGAGATGAAGAAATCTTGAAAGTTTCCTTAGATCAATTA 1083
QY 155 SerValValAlaThrSerLeuSerHis-----PheIleProSerValIleSerThrAsn 173
DB 1084 AATATGTTAAAGATTAATTAATTAATGAATTAATTAATTAATTAATTAATTAATTA 1143
QY 174 -----LysAspIleLeuHisAsnValAlaGlnIleLeuHis-----CysTyr 187
DB 1144 GTTTAAATTAATTAATTAATTAATTTATTATTAAGATTTATTATTAACAACATATGATCAT 1203
QY 188 AlaGluuysLeuIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleVal 207
DB 1204 AAGAAAGAAAGAAAGCTCAAGAA-----AAGGAATTACCA----- 1236
QY 208 TyrArgArgPheGlnProAlaLeuIleAspValPheGlnValLeuGluSerAspPro 227
DB 1237 -----GAACCTACTGTACTAAT----- 1254

QY 228 GluysPheProThrGluLeuLysLeuIleGluSerAsnTrpGluGluuysHisGlu 247
DB 1255 GAAGAATATGTTGAAGAAATTAAGAAAGTAT----- 1287
QY 248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlySerGlu 267
DB 1288 -----CTAGATATGGGTATCAAAATTAATTAATTAATTAATTAATTAATTAATTA 1332
QY 266 ThrProLysLeuValSerArgValLeuGlnTrpLysGlu-----LysPro 281
DB 1333 TTAATAAATTAATAAATAATTAATTTCCCTTAAGAAAGAAAGAAATTAATCAAGCAGTAGAT 1392
QY 282 ----- 283
DB 1393 ACCAAAGATGAGAAAGAAAGCAAGTAAAGCAACAGCTTAAAGAGTGTAAACCA 1452
QY 284 GluGluSerSerValValTyrAspGlnLeuHisSerAlaAsnLeuGln-----Phe 300
DB 1453 ACGGAAGATTTCTTAATTAATTAAGAACAGTATTAATTAATTAATTAATTAATTAATTAAT 1512
QY 301 MetLysGluLeuArgGlu----- 306
DB 1513 GAAGAAGATTAATCAAAATTAATTAATTAATCACTAATGTTTACCACTCAATCAAAA 1572
QY 307 -----MetArgLysTyrAspSerAspProGlu 316
DB 1573 AAAAAAACAATAAGAACTGATCTGATAGTAAATTTTGAATATCATCTGAA 1632
QY 317 ThrTyrIleLysGlu-----LeuAspHisSerValGlu 327
DB 1633 AATTATTTTAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1692
QY 328 ProLeuThrValAlaIleLys-----AsnIleArgLysGly 339
DB 1693 AAAATAGGTGTCACTTAATAAATAATTAAGACCTTAATAAATGAAGAAAGTTATGTAAC 1752
QY 340 LeuGlnAlaLeu-----ThrGlnLysSerGluValProIleGluProAsp 354
DB 1753 ATTAATGATTCATTAGAAATTAAGATTAATAAACAATTGAAGCATTAACAACAT 1812
QY 355 ValGlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValAla 374
DB 1813 ATTCAAAATTAATTA-----CAAGATTTAACAAGCTTAATTAATTAATTAATTAAT 1860
QY 375 ProGlyAlaGlyGlyTyrAspAlaIleAlaValLeuValLeuGluAsnGlnValGlyAsn 394
DB 1861 AATTATTAACAATGGAACAATAATTAATTAACAATA----- 1896
QY 395 PheLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyrTrpValAspLeu 414
DB 1897 TTTCACAAATAATTAATCTGAATAAT-----GATGTTCTTAATCAAGAAACGAGAGAAAGATG 1953
QY 415 GluGluGlnThrGluGlyValLeuGluGlu 424
DB 1954 GAAAAACAAGTTGAAGCAATCAACAAGCA 1983
RESULT 7
US-09-595-684B-30
Sequence 30, Application US/09595684B
Patent No. 6544766
GENERAL INFORMATION:
APPLICANT: Beraud, Christophe
APPLICANT: Ohashi, Cara
APPLICANT: Sakowicz, Roman
APPLICANT: Vaisberg, Eugeni
APPLICANT: Wood, Kenneth
APPLICANT: Yu, Ming
TITLE OF INVENTION: Human kinesins and methods of producing
TITLE OF INVENTION: And purifying human kinesins
FILE REFERENCE: cytopo36
CURRENT APPLICATION NUMBER: US/09/595,684B
CURRENT FILING DATE: 2002-06-24
PRIOR APPLICATION NUMBER: 09/295,612

PRIOR FILING DATE: 2000-04-20
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 30
LENGTH: 8257
TYPE: DNA
ORGANISM: Human
US-09-595-684B-30

Alignment Scores:

Score:	108.50	Length:	8257
Percent Similarity:	37.28%	Matches:	67
Best Local Similarity:	20.74%	Conservative:	84
Query Match:	4.87%	Mismatches:	161
DB:	4	Indels:	93
		Gaps:	16

US-10-069-062-7 (1-432) x US-09-595-684B-30 (1-8257)

```
QY 47 Iysgluserargileysileserseproglinphelaasnglyglutprglutyrhis 66
DB 2644 AAGAGAGCCCAAAATTTGATTCGAGTTGGGTGCTTGAAGACGAGCTTCTTACAG 2703
QY 67 IleserSersanthglulysproargluvalglinserygileasnprophleuglu 86
DB 2704 ACCCAAGAACTTCAGAGAAAACCGTGTGCTTCAAGAAAGACTAATGAG--ATGGA 2760
QY 87 Alathrilepheilevalleualaryrileglinprothrgluvalpheaspheuglu 106
DB 2761 CAGCGAAGAAACAATTAAATAATGAGATTCCTCGCTGCAACCTGTAAGAAAGGAGAA 2820
QY 107 ileiletyrseraspProglityrhisserglingluasprthrglutrllyser 126
DB 2821 ACACGTGATTACTGAGAACTGACGAAACTTGAAGAAAGTAACCTTAACTCA 2877
QY 127 Aenglyglulysrthpheleutyrrhisserargalallethrgluvalglulysrthgly 146
DB 2878 -----GAAAAGATGATCTTAAACAACCTCCAAAGAAAGCTTCAATGAGAGGACCA 2931
QY 147 LeuglyserSersala----- 151
DB 2932 CTCGAAAGTGAATTCACGATCTGTTAACATGAATATGATATCTCAAGAACATTA 2991
QY 152 ---glyleuvalserVal-----ValalathrserleuSerhisshelle 166
DB 2992 AATGCTCTGAGTCTGTGAACAACATCAAGAAACAATTAATACCTAAATCGAAAT 3051
QY 167 ProasnvalIleserThrasnlyaspilleuhsasnvalalaglullealhisCys 186
DB 3052 TCTGAGGAATTTCCAGGAATTTGCATATGAGAAATAACAGAGAAACTAAAGATGA 3111
QY 187 TyrAlaglulysleilegly-----Serglypheasp 197
DB 3112 TTTCAAGCAAAAGATGTTGGCATGATATAAAAAACGATTGGAGCTAAAAATACCA 3171
QY 198 ValalathrVala-----iletyrgly 204
DB 3172 ACACTAACCTGCAGAGATTAAAGATAATGATTAAGCAACAAGAAAGATATTTCT 3231
QY 205 LeuilevalIlyrargargpheglinProalaleuileasnspvalpheglinvalleuglu 224
DB 3232 TTAATACAGAGAAA-----AATGAACCTCAACAATGTTAAG 3270
QY 225 Ser-----AspProglulyspheProthrgluLeuileuvalleuilegluser 241
DB 3271 ACTGTTATACGCAAGAAAGCAATTTGAAGACTTAAGCAAAATTTGAATAATGAC 3330
QY 242 TrpglugluysrhislegluargCysrthleuprotyrglyleleuLeuileuMetCglyasp 261
DB 3331 ATTGAAGAACCAAGAGAA-----TTAAGACTTCTT----- 3360
QY 262 VallysglyglyserglutrProlyserleuValSerargvalleuglntprlysls--- 280
```

```
DB 3361 -----GGGAGTAACTTAAAAACAACAGAGATAGTTGCAACAAGAAAGACCAT 3411
QY 281 -----GluysProglulyserserValIlyrAspSlnleuasnseralasnleu 298
DB 3412 GCCATTAAGAAAGAAAGAGAGAGCTTTTAGACCTGTGACAGCTGCGCAAGCTTGAAGA 3471
QY 299 GlnpheMetlysgluLeuarggluwerargglulysrtyrAspSer-----AspProglu 316
DB 3472 AAACCTAAAGAAAGAAAGACGACGACACTCCAGAAAAAGCAACAACCTCTTAATGACAA 3531
QY 317 ThrtyrilelysgluLeuAspHisserValgluproleuthrValalalleysanlle 336
DB 3532 GAGAGAGAGTGTGATGTCAGAAAAAGATTAT-----GAATATGAGATTTA 3579
QY 337 ArglysglyleuglinalleuthrglulyssergluvalProilegluProaspValgin 356
DB 3580 AAGAAATGAATTA-----AAGAACAAAGAAATGACATTGAA--CATATGGA 3624
QY 357 ThrGlnleuLeuAspArgCysglnglutlileProglyCysValglyValalProgly 376
DB 3625 ACAGAGAGCTTGAAGTGTGCTCAGAACTT----- 3654
QY 377 AlaelyglytyrAspAlallealValleuValleugluasngluvalglyAsnphelys 396
DB 3655 AATGAAATTTATGAGAAAGTGAATCTATAACCAAGAAAGAAAGTTCTAAAGGAATTA 3714
QY 397 GlnysThrleuglu 401
DB 3715 CAGAGTCATTGAA 3729
```

RESULT 8

US-09-620-312D-130

Sequence 130. Application US/09620312D

Patent No. 6569662

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod

APPLICANT: Zhang, Jie

APPLICANT: Ren, Feiyun

APPLICANT: Chen, Rui-hong

APPLICANT: Zhao, Qing A.

APPLICANT: Weinman, Tom

APPLICANT: Xue, Aidong J.

APPLICANT: Yang, Yonghong

APPLICANT: Wang, Jian-Rui

APPLICANT: Zhou, Ping

APPLICANT: Ma, Yungqing

APPLICANT: Wang, Dunrui

APPLICANT: Wang, Zhilwei

APPLICANT: John Tillinphast

APPLICANT: Drmanac, Radoje T.

TITLE OF INVENTION: No. 6569662el Nucleic Acids and

FILE REFERENCE: 784CIP2B

CURRENT APPLICATION NUMBER: US/09/620,312D

PRIOR FILING DATE: 2000-07-19

PRIOR APPLICATION NUMBER: 09/552,317

PRIOR FILING DATE: 2000-04-25

PRIOR APPLICATION NUMBER: 09/488,725

NUMBER OF SEQ ID NOS: 1105

SOFTWARE: DE_FL_genes Version 1.0

SEQ ID NO 130

LENGTH: 8503

TYPE: DNA

ORGANISM: Homo sapiens

FEATURES:

NAME/KEY: CDS

LOCATION: (91)..(8082)

US-09-620-312D-130

Alignment Scores:

Pred. No.: 0.104 Length: 8503
 Score: 108.50 Matches: 84
 Percent Similarity: 37.28% Conservative: 67
 Best Local Similarity: 20.74% Mismatches: 161
 Query Match: 4.87% Indels: 93
 DB: 4 Gaps: 16

US-10-069-062-7 (1-432) x US-09-620-312D-130 (1-8503)

QY 47 LysGluSerArgIleLysIleSerSerProGlnPheAlaSerGluIleProGluIlePhe 66
 DB 2644 AAGAGAGCCCAAAATTTGATTCGAGTTGGGTCCTTGAAGCCAGCTTCTTCAAG 2703
 QY 67 ILeSerSerAsnThrGluIleProArgGluValGlnSerArgIleAsnProPheLeuGlu 86
 DB 2704 ACCCAAGACTTCAG 2760
 QY 87 AlaThrIlePheIleValLeuAlaIleIleGlnProThrGluAlaPheAspLeuGluIle 106
 DB 2761 CAGCTGAG 2820
 QY 107 IleIleIleSerSerProGluIlePheIleSerGlnGluAspThrGluIlePheSerSer 126
 DB 2821 AACCTATTCAG 2877
 QY 127 AsnGluIleIleThrPheLeuIleIleSerArgAlaIleThrGluValGluIlePheThrGly 146
 DB 2878 GAAAAAGATGATCTAAAG 2931
 QY 147 LeuGluSerSerAla----- 151
 DB 2932 CTCGAAAGAGATGATTCAG 2991
 QY 152 GLeuValSerVal-----ValAlaThrSerLeuLeuSerHisPheIle 166
 DB 2992 AATGCTCTTGAGTCTCTGAAACAACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3051
 QY 167 ProAsnValIleSerThrAsnIleLeuIleAsnValAlaGlnIleAlaHisCys 186
 DB 3052 TCTGAG 3111
 QY 187 TyrAlaGlnIleLysIleGly-----SerGlyPheAsp 197
 DB 3112 TTTTCAGCAAAAGATGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 3171
 QY 198 ValAlaThrAla-----IleTyrGly 204
 DB 3172 AACCTATTCAG 3231
 QY 205 LeuIleValIleTyrArgIlePheGlnProAlaLeuIleAsnAspValPheGlnValLeuGlu 224
 DB 3232 TTTATTCAG 3270
 QY 225 Ser-----AspProGluIlePheProThrGlnLeuIleLysIleGluSerAsn 241
 DB 3271 AGTGTATATGAG 3330
 QY 242 TTPGluGlnIleHisIleGluIleCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAsp 261
 DB 3331 ATTGAAG 3360
 QY 262 ValIleGlyIleSerGluIlePheProThrGlnLeuValIleSerArgValLeuGlnIlePheLys 280
 DB 3361 GGGATATGAG 3411
 QY 281 GlnIleProGluIleGluIleSerValIleTyrAspGlnLeuAsnSerAlaLeuLeu 298
 DB 3412 GCCATTAAG 3471
 QY 299 GlnPheMetIleGluIleLeuIleArgIleMetIleArgIleIleTyrAspSer-----AspProGlu 316
 DB 3472 AAATTAAG 3531

QY 317 ThrTyrIleLysGluLeuAspHisSerValGluProLeuThrValAlaIleLysAsnIle 336
 DB 3532 GAAGAGATGAG 3579
 QY 337 ArgIleGlyLeuGlnAlaLeuThrGlnIleSerGluValProIleGluProAspValGln 356
 DB 3580 AAGATGAGATTA-----AAG 3624
 QY 357 ThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValProGly 376
 DB 3625 AAG 3654
 QY 377 AlaGlyGlyTyrAspAlaIleAlaValLeuValLeuGluIleAsnGlnValGlyAsnPheLys 396
 DB 3655 AATGAAGATTAAG 3714
 QY 397 GlnIlePheLeuGlu 401
 DB 3715 CAGAGATCTTTGAA 3729

RESULT 9
 US-09-198-452A-1
 Sequence 1, Application US/09198452A
 Patent No. 6559294
 GENERAL INFORMATION:
 APPLICANT: Grifflair, R.
 TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
 TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
 FILE REFERENCE: 9710-003-999
 CURRENT APPLICATION NUMBER: US/09198,452A
 CURRENT FILING DATE: 1998-11-24
 NUMBER OF SEQ ID NOS: 6849
 SEQ ID NO 1
 LENGTH: 1230025
 TYPE: DNA
 ORGANISM: Chlamydia pneumoniae
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: (1) - (15000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc.feature
 LOCATION: (15001) - (30000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc.feature
 LOCATION: (30001) - (45000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc.feature
 LOCATION: (45001) - (60000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc.feature
 LOCATION: (60001) - (75000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc.feature
 LOCATION: (75001) - (90000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc.feature
 LOCATION: (90001) - (105000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc.feature
 LOCATION: (105001) - (120000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc.feature
 LOCATION: (120001) - (135000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc.feature
 LOCATION: (135001) - (150000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc.feature
 LOCATION: (150001) - (165000)
 OTHER INFORMATION: n=a or c or g or t
 NAME/KEY: misc.feature
 LOCATION: (165001) - (180000)


```

; LOCATION: (900001)..(91500)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature

Alignment Scores:
Pred. No.: 426 Length: 123025
Score: 108.00 Matches: 75
Percent Similarity: 37.80% Conservative: 69
Best Local Similarity: 19.69% Mismatches: 133
Query Match: 4.84% Indels: 104
DB: 4 Gaps: 18

US-10-069-062-7 (1-432) x US-09-198-452A-1 (1-1230025)

QY 36 HisAlaValIleThrProLysGlyThrSerLeuLysGlySerArgIleLysIleSer 55
DB 693349 TATGTCGCTTACG-----GGAAATCTCGCGCAACCTTACCAATCAAGCTT----- 693396
QY 56 ProGlnPheAlaAsnGlyLutRgIutRgIutRgIutRgIutRgIutRgIutRg 75
DB 693397 -----CAGGAATCTCTCAAGCCCTCAAGCAAAAGCC----- 693429
QY 76 GluValGlnSerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTy 95
DB 693430 -----ATAATGCTATGTCGATATTAACAATTAACATGCTTCTCTCA 693474
QY 96 IleGlnProThrGluAlaPheAspLeuGluIleIleIleTySerAspProGlyTyRHis 115
DB 693475 GGGCAACCTCTGACGCTTACGATGCGAGTCAGCTGCTTACCTCTCGCAAGTAA 693534
QY 116 SerGlnLysPThrGluThrIleThrSerAsnGlyLysThrPheLeuTyRHis 135
DB 693535 AAGCTCCACCCCAAGATCTCAACCTATTAAGAGAAACGTCCTTGCCCTCA 693594
QY 136 SerArgAlaIleThrGluValGlyLysThr-----GlyLeuGlySerLeuGly 152
DB 693595 GGAGTCCCTGATCGCGCATGATGATCTCTGAGTCTTATGAGGCG 693654
QY 153 -----LeuValSerValAlaThrSerLeu-----SerHisPheIle 166
DB 693655 AAGGACCCCTCATTTCAAGAAACCAACACCACTACAGTCATCAAGCTGCTATTCCTC 693714
QY 167 ProAsnValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCys 186
DB 693715 CCCGAAGCTCTCGGCTCTCAAAACTCTCCGATTCGATCGGAATGCTGCTTACGA 693774
QY 187 TyrAlaGlnLysIleGlySerGlyPheAspValAlaThrAlaIleTyRLeuIle 206
DB 693775 TTCACCCCG-----GGGATCGATCCCAAAATGTTGTA----- 693807
QY 207 ValTyArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGlySerAsp 226
DB 693808 -----CCAGACTACCAAGCTCAATTCATATTTTAAAGATCTTC 693849
QY 227 ProGlnLysPhe-----CCAGACTACCAAGCTCAATTCATATTTTAAAGATCTTC 230
DB 693850 CCCGAAGTACAAATCTCCCATCTATGTTCTGAGAAATTTGCTGTAATTAAGAG 693909
QY 231 -----ProThrGluLeuLysLysLeuIleGlySerAsnTrpGluLysHis 246
DB 693910 GTCGCTCTACGCCCTAAACCCCTACAGAAATCTGAGGAAATCT----- 693954
QY 247 GluArgCysThrLeuProTyRgIyIleLysLeuLeuMetGlyAspValLys-----Gly 264
DB 693955 -----TTCCTCAATGAGATCTCTCTCAAAAGTTACAGAGCTTAAAGG 693996
QY 265 GlySerGluThrProLys-----LeuValSerArgValLeuGlnTrpLysGlu 281
DB 693997 TTCTTACAGACTCACAGAACTTCTTAAAGTCCCTTCTCAACGCGCATGAC 694056
QY 282 LysProGlnGluSerSerValValTyRAsp-----GlnLeuAsnSerAlaAsnLeuGln 299
DB 694057 ATCAATGAAGAATAATAGATCTAGTAAAGATCTGTAGGACGAAATCTTGAAATATAGAA 694116

```

```

QY 300 PheMetLysGluLeuArgGluMetArgGluLysTyRAspSerAspProGluThr----- 317
DB 694117 ACTCAAAATCCAGTATCTCTGCTACACTCCAAATCTCAAAAGTAAAGTAACTCTGGG 694176
QY 318 -----TyrIleLysGluLeuAspHisSerValGlu 327
DB 694177 TTCTTAGCAAAACGACGACTTCAGAAATCTTCACTCTGACCTG-----CTAGAT 694227
QY 328 ProLeuThrValAlaIleLysAsnIleArgLysGlyLeuGlnAlaLeuThrGlnLysSer 347
DB 694228 CCCGAAACAGTGGCTCTA-----ACAAGAAA-----GAAAGAGA 694263
QY 348 ProIleGlnProAspValGlnThrGlnLeuLeuAspArgCysGlnGluIlePro 367
DB 694264 GAAATCTCTCTCAGGCTCCCAAAATACCACTGATTC--AGATCTCACTGCTTCA 694320
QY 368 Gly 368
DB 694321 GGA 694323

RESULT 10
US-09-107-532A-208
; Sequence 208, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSER: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 208:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2451 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...2451
; SEQUENCE DESCRIPTION: SEQ ID NO: 208:
US-09-107-532A-208

```



```

QY 63 TRPGLUTYRHISLEISERSEANTHR-----GLUTYPEPROARGGLVALGINSERARG 80
DB 6480 TGGAAATGCGATCTCCGTGGAAATCAATGATGATTCCTTAAATGAACAGATGGCCAA 6421
QY 81 ILEASNPROPHLEUGLUALATHRIIEPHEILEVALLEUALATYRIIEGINPROTHRGU 100
DB 6420 TTTAAACGGAATTTGGAAGGAAGTAGCTGATGATTCAGTATTTCAACT----- 6367
QY 101 ALAPHAPELEUGLULEILEILETYRSEAPPROGLTYRTHISERGINGLUAPTNR 120
DB 6366 -----CAGGCGGACCGACCTATTGCGGAAAAATTTTCA 6331
QY 121 GLUTHRYSETHSERSEANGLYGLUTYRTHPELEUTHYRTHISERAGLAIETNR 140
DB 6330 AATTATTAACCGTCTGCAATTCGATCGCACTTAT-----CGGAATTGGAG 6283
QY 141 GLUVALGLULYRTHRGYLEUGLYSERSERALGLYLEUVALSERVALALATHRSER 160
DB 6282 GAGATGGCCAAAG-----CAAGTGACCAATCCTTCT 6253
QY 161 LEULESERHISPEHEILEPROAENVALIIESETHRASNLYAPRIIELEUHIISANVAL 180
DB 6252 TCTCGTTATTATCAAGATGAACCTGCTCGAAGCACTGAGGATTCATGAGATGAGT 6193
QY 181 ALAAGLILEALAHISCYSTRALAGLULYLYS-----ILEGLYSERGYPHEAPVAL 198
DB 6192 CATTAAG-----CATGCTTACATAGTGAAGAAAGAGATTTGCGAAGCTGGCGAT 6142
QY 199 ALATHRALIETRYGLYLEUILEVALTYRARGYPHEGINPROALALEUANAP 218
DB 6141 -----TATGAATCGGTACACCTGCTGCTCAACCAAT 6109
QY 219 VALPHEGLINLEUGLU-----SERAPPROGLU-----LYSPHEPROTHRGU 233
DB 6108 ACCCTGCTCTGATGAAGAAATCTTCTGATGAGAAATTAATAATATACAGATGTG 6049
QY 234 LEULYALYLEULEIEGLUSEAENTRGLUGLULYHISGLULYRGYSTRILEUPROYR 253
DB 6048 ATTGAABAAATTGTACCA-----GATCCGGAACATTTCCGAAGACGACTGATAC 5998
QY 254 GLYLILELYLEULEUMETGLYAPVALYSGLYSGLYSERGLUTHRPROLYSEUVALSER 273
DB 5997 CCATTCAGAGCTCTAGGTGAACCTAGTATGAGGAAGGTA---AAAGTAATAGCT 5941
QY 274 ARGVALLEUGLINTRPLYSGLULYSPROGLUGLUSESERVALVALTYRARGINLEU 293
DB 5940 GGTTCAGT-----CGTAAGAGATCAAGAAATTTCTTCACT 5899
QY 294 ASNSEKALASNLEUGLNPHEMETLYSGLEUARGGLUMETARGGLULYSTRAPSER 313
DB 5898 -----CGCTCGATGAGCAAGTTCACACTTGTGACCAAGGTGAAGTTTTCACA 5845
QY 314 AAPPROGLUTHRYRIIELYSGLEUAPRHSERVALGLUPROLEUTHRALALAIIE 333
DB 5844 GAT---GGAATCTATATC-----GACCAACCAAGTTGCTTATAGCGTCTTAT 5797
QY 334 LYSANLLEARG---LYSGLYLEUGLINALALEU-----THRGILYSESERGLUVAL 349
DB 5796 GGGAAATGTTTGTGATGAGCGCTGTCTCAACTGTTCACACTTCAAAAGACCAAGAAAT 5737
QY 350 PROILEGLUPROAP----- 354
DB 5736 CCAATGATTAATAAATGCAACACATGTACACTGATGATAATCGTTTGTCTCT 5677
QY 355 -----VALGINTHRGINLEUAPARGYSGINGLULEPROGLYCYSEVALGLYGLY 372
DB 5676 TTGCTGTGATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5635
QY 373 VALVALPROGLYALA-----GLYGLYTYRASPALALEALALEU----- 386
DB 5634 TCGATCAGTCGTCAAAATAGCGAGGGGCAAGTGGCCGACAGTGAAGTACTTAAGAGGAT 5575

```

```

QY 387 -----VALLEUGLUSNGLINVALGLYASNPHELYSGIN-----LYS 398
DB 5574 CACCGAATGCGGATATGCTGATGAGGAAGAAACCAACATGTTTGCAGTGTGGAAG 5515
QY 399 THRLEUGLUSNPROAPPTYPHEHISANVALTYR 410
DB 5514 ACCATGTTCAATCGATGATGATATGATGATGATGATGATGATGATGATGATGAT 5479
RESULT 12
US-09-484-970B-60
/ Sequence 60, Application US/09484970B
/ Patent No. 6426186
/ GENERAL INFORMATION:
/ APPLICANT: Jones, Karen A.
/ APPLICANT: Volkmuth, Wayne
/ APPLICANT: Walker, Michael G.
/ TITLE OF INVENTION: BONE REMODELING GENES
/ FILE REFERENCE: PB-0014 US
/ CURRENT APPLICATION NUMBER: US/09/484,970B
/ NUMBER OF SEQ ID NOS: 172
/ SOFTWARE: PERL Program
/ SEQ ID NO 60
/ LENGTH: 13977
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. 6426186 229357.11CB1
/ NAME/KEY: unsure
/ LOCATION: 11721-11761, 12294, 13969
/ OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-60

```

```

Alignment Scores:
Pred. No.: 1.86 Length: 13977
Score: 101.00 Matches: 95
Percent Similarity: 38.38% Conservative: 80
Best Local Similarity: 20.83% Mismatches: 136
Query Match: 4.53% Indels: 148
DB: Gaps: 21

```

```

US-10-069-062-7 (1-432) x US-09-484-970B-60 (1-13977)
QY 56 PROGINPHEALASNGLYGLUTHRPLGLUTYRHISLEISERSEANTHRGLULYSPROARG 75
DB 6030 CCGCAATG-----CAGTGAAGCCAACTCAAGTCCAGTCC-AGCAAGCGCTGGCG 6079
QY 76 GLUVALGINSERARG-----ILEASNPROPE 84
DB 6080 GAAATTTGAGCAAAATTTGCTGATTCGAAAGACTCAACTTTCACAAATTCACACTGTC 6139
QY 85 LEUGLUALATHRIIEPHEILEVALLEUALATYRIIEGINPROTHRGILUALAPHEAPLEU 104
DB 6140 CGTGAAGAAACATATGATGATG-----ACTGAAGACATGCTTTG 6181
QY 105 GILILEILEIETYSERAPPROGLTYRTHISERGINGLUAPTNRGLUTHRYSTR 124
DB 6182 GAAATTTCTTATGTC-----CTTCTACTTATTTGATCGAAATCACTCATGTC----- 6225
QY 125 SERSEANGLYGLULYRTHPELEUTHYRTHISERAGLAIETNRGLUVALGLULY 144
DB 6230 -----TACAAGCCCTATTGAAGTGAACAA 6256
QY 145 THRGLYLEUGLYSERSERALAGLYLEUVAL-SERVALVALALATHRSERLEULESERHI 164
DB 6257 -----CTTCAATGCTCTGACCTGTGCTTACAGAC 6289
QY 164 SPEHEILEPROAENVALIIESETHRASNLYS-----ASPILELE 177
DB 6290 TTG-----AAGATCTCTTAAAGCAAGAGAGCTCTGGAAGATTAATAAGATAT 6340
QY 177 WHISANVALALAGLINALALEA-----HISCYSTRY 187

```



```

Db      6341 CTACCAAGCTCAGGTCGATTGACATTATTCATGACAGAGACAGACATTGCA 6400
Qy      187 rAlaGlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleVa 207
Db      6401 AGTGCACGCC-----CTGTGAAAAGGCTGAAG 6427
Qy      207 lTyrArgArgPheGlnProAlaLeuIleAsnAsp----- 218
Db      6428 CTACAGGAAGCTCTCTCCACGCTGATTTCCATGAGAAAAGTTAACAAAATGTACAG 6487
Qy      219 -----ValPhe----- 220
Db      6488 GACCGACAGGCGATTGACAGATCTGTGAGAAATGCGCGCTTTTCATTATGATATA 6547
Qy      221 -GlnValLeuGlnUserAspProGlnLysPheProThrGlnLeuLysValLeu---lGql 239
Db      6548 AAGATATTTATCATGAGCTTAACGAAAGCTGAACAGTTT-CTCAGAAAGACACAAATTC 6606
Qy      239 uSerAsnTrpGlnGluLysHisGlnArgCysThrLeuProTyrGlyIleLysLeuLeu 259
Db      6607 TGAGATTGGGAA-----CATGCTAA-----TACAAATGGTATCT 6642
Qy      259 rGlyAspValLysGlyGlySerGlnThrProLysLeuValSerArgValLeuGlnTrp 279
Db      6643 TAGAGAACTCCAGATGCGATTGGCGACGCGCAACTGTTCAGAAACATTGAATGCAAC 6702
Qy      279 eLysGlnLys-----ProGlnGlnUserSerValValTyrAspGlnLe 293
Db      6703 TGGGAGAGAAATTAATTCAGCAATCTCTCAAAAACAGATGCCAGATTTCTACAGAAAAT 6762
Qy      293 uAsnSerAlaAsnLeuGlnPheMetLysGlnLeuArgGlnMet-----ArgGlnLysTyr 311
Db      6763 GCGAAGCTCGATCTGCGGTGGCAGAGAGCTTCGCAACAGCTGTCAACAGAAAAAGG 6822
Qy      311 rAspSerAspProGlnThrTyrIleLysGlnLeuAspHisSerValGlnProLeuThrVa 331
Db      6823 GCTAGAGAGAAAGAAATATCTGTTCAGAAATTTCAAGAGATTTAAATGATTTGTTT 6882
Qy      331 lAlaIleLysAsnIleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGlnValPro 351
Db      6883 A-----TGGTTGAGAGACAGACATTAACATTGACGATCCACT 6921
Qy      351 eGlnProAspValGlnThrGlnLeuLysAspArgCysGlnGlnIleProGlyCysValG 371
Db      6922 TGAACCTGGAAGAGACGACGACCTAAAGAAAGCTTGAACAGCACTT-ACGTGTGG 6980
Qy      371 yGlyValValProGlnAlaGlyGlyTyrAspAlaIle-----AlaVal-Leu 387
Db      6981 AAGAGTTCCTCCGCGCCAGCAATTCCAACATTAATGAACCTGAGAGACCCGTGC 7040
Qy      387 AlLeuGlnAsnGlnValGlnAsnPheLysGlnLysThrLeuGlnAsnProAspTyrPhe 407
Db      7041 TTGTAGTGTCTCCCATAGCCCAAGAGCAAGATTAACCTGAAAATAGCTCAAGCAGA 7100
Qy      407 lAsnValTyrTrpValAspLeu-----GlnGlnGlnThrGln----- 419
Db      7101 CAATCTCAGTGTGAATTAAGCTTCCAGAGCTTTACTGAGAAACAGAGAAATTTGAAG 7160
Qy      420 -----GlyValLeuGlnGlnLysProGlnLys 428
Db      7161 CTCAAAATTAAGACCTTGGGCGAGCTTGAAAAAAAGCTTGAAGAC 7204

```

RESULT 13
US-09-251-645-11

Sequence 11, Application US/09251645

Patent No. 6281413

GENERAL INFORMATION:

APPLICANT: Kramer, Vance C.

APPLICANT: Morgan, Michael K.

APPLICANT: Anderson, Arne R.

APPLICANT: Hart, Hope

APPLICANT: Warren, Gregory W.

```

/ APPLICANT: Dunn, Martha
/ APPLICANT: Chen, Jeng S.
/ TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
/ TITLE OF INVENTION: AND NUCLEIC ACID SEQUENCES CODING THEREFOR
/ FILE REFERENCE: CG1963/A
/ CURRENT APPLICATION NUMBER: US/09/251,645
/ NUMBER OF SEQ ID NOS: 22
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 11
/ LENGTH: 37948
/ TYPE: DNA
/ ORGANISM: Photorhabdus luminescens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (15171)..(18035)
/ OTHER INFORMATION: orf5
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (23768)..(31336)
/ OTHER INFORMATION: hph2
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (31393)..(35838)
/ OTHER INFORMATION: orf2
/ US-09-251-645-11

```

Alignment Scores:

```

Pred. No.: 9.63 Length: 37948
Score: 101.00 Matches: 103
Percent Similarity: 32.89% Conservative: 70
Best Local Similarity: 19.58% Mismatches: 193
Query Match: 4.53% Indels: 160
DB: 3 Gaps: 24

```

US-10-069-062-7 (1-432) x US-09-251-645-11 (1-37948)

```

Qy      19 ValLeuGlnProIleTyrAspAlaTyrValThrAlaLeuSerSerArg----- 34
Db      3718 ATACTTCTCCAGAAAGAAATAGCAATTACTGATCTGATCGCGATACACTAGTACA 3777
Qy      35 MetHisAlaValIleThrProLysGlyThrSerLeuLysGlnSerArgIleLysIleSer 54
Db      3778 CAACAGACTATCGAAATCTGATTAATGCTGCTCA-----TTGAAAATTAAT 3825
Qy      55 -----SerProGlnPheAlaAsnGlyGlnTrp 63
Db      3826 AAGTTATCTGTATTCAAAGCTATAAATTATCACCGACGATTAATGAATGAAATCCTC 3885
Qy      64 GluTyrHisIleSerSerAsnThrGlnLysProArgIleValGlnSerArgIleAsnPro 83
Db      3886 CGC-----AGCATCAGCAGAA-----TTCAATATCAATATAA 3918
Qy      84 PheLeuGlnAlaThrIlePheIleValLeuAlaTyrIleGlnProThrGlnAlaPheAsp 103
Db      3919 GAATATATTAACAAAATCTTTCGTTAATATATATATGAAA-----CGTTATGATTTGAC 3975
Qy      104 LeuGlnIleIleIle-----TyrSerAspProGlyTyr 114
Db      3976 ACTAGACTGCATTAATATATGACAGGTACCAATTTCACATATATCAATGAACGACAT 4035
Qy      115 HisSerGlnGlnAspThrGlnLysThr-----SerSer 126
Db      4036 CTAACTCAGTTGATCGTTTATTAATTTCCCAACTGAATGCTCAAGATTTTCCGTC 4095
Qy      127 AsnGlyGlnLysThrPheLeuTyrHisSerArgAlaIleThrGlnValGlnLysThrGly 146
Db      4096 AATGTACTCAGATTAATTAATTAACCTTAAGCAGTACCAACACTGGAATTAACAGTA 4155
Qy      147 LeuGlySerSerAlaGlyLeuValSerValAlaIleThrSerLeuLeuSerHisPheIle 166
Db      4156 CTTAACGTCCTTTAACTCGATGATATTCATTAATGACACTATAAA-----ATT 4209

```

```

QY 167 ProAsnValIleSerThrAsnValLeuHisAsnValAlaGlnIleAlaHisCys 186
DB 4210 ACCAAATCCGGTCAATCTACCGAAAGATATACATATATAGATATCTTCTCATTC 4269
QY 187 TyrAlaGlnLysLeuIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIle 206
DB 4270 TATAGACAAATTCATCGACAGATATCCATCATGATTACATGATGATGACGGGTACAG 4329
QY 207 ValTyrArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGlySerAsp 226
DB 4330 TTG-----GAGCCATAGGTAAAGAAACAACCAATTAATCTGATTACT 4374
QY 227 ProGluLysPheProThrGluLeuLysLeuIleGlu-----SerAsnTyr----- 242
DB 4375 CCTGACAAATCTGGTTACTCTAATTAACAACCTATGCTGCTACGTCAGTGGCTACGTACA 4434
QY 243 -----GluGluLysHisGluLysArgCysThr 250
DB 4435 CAAAGTGAAGTGTATCATGTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 4491
QY 251 LeuProTyrGlyTyrLeuLysLeuMetGlyAspValLysGly----- 265
DB 4492 CTAAACCCGGAATTCGAAATTAATCTGATACCGCTTACATGCTTACAGATTTTAAT 4551
QY 266 -----SerGluThrProLysLeuValSerArgValLeuGlnTyrLys 279
DB 4552 AAGAAGATGTGAAGATGAAAGATCTAGAGAGAAACCAAAAGAAATTCGAGAGCCG 4611
QY 280 LysGluLys-----ProGluLysSerSerValTyrArgGlnLeuAsnSer 295
DB 4612 AAGGAAATCTGGAAATTTCCCGGAAACCAAGCAACACTCCAGAAAGACAGAAATAA 4671
QY 296 AlaAsnLeuGlnPheMetLysGluLeuArgGluMetArgGlnLysTyrAspSerAspPro 315
DB 4672 GCCCAGAGAGA-----CTGAATTAAGGTGAAGAGACCCAGAAACCAAGAAACCTTA 4728
QY 316 GluThrTyrIleLysGluLeuAspHisSerValGluProLeuThrValAla----- 332
DB 4729 GATGAGTCAAGAAATCTGCAAAAGCATATCTCTTATATGCCCCGCTGCA 4788
QY 333 -----IleLysAsnIleArgLysGly 339
DB 4789 TTACCATCTGAACATCGGCATATTCATATCTCATCTGGGAGATATCTGGAAACCCGCG 4848
QY 340 LeuGlnAlaLeuThr-----GlnLysSerGluValProIle 351
DB 4849 ATAGGAAATATGACAGCGGAAATTAATGGAACCTGCTGGGAAATCCGTTACGCGCT 4908
QY 352 GluProAspValGlnThrGln-----LeuLeuAspArgCysGlnGluIleProGly 368
DB 4909 CAACCTGAATTCAAAAACAAGCTGAACCTGTGCTCAGATTTGCCAGCGC----- 4959
QY 369 CysValGlyValValProGlyAlaGlyTyrAspAlaIleAlaValLeuValLeu 388
DB 4960 -----CTGGCACAACACTAGCGTGG 4977
QY 389 GluAsnGlnValGlyAsnPheLysGlnLysThrLeuGlu-----AsnProAsp 404
DB 4978 ATTAAACGTTCTACCGGCTTAAAGAAACACCTTAAGTCTGTTGTGCAAAAGCGGCA 5037
QY 405 TyrPhe-----HisAsnVal 409
DB 5038 CACTTTGTTATTAATAACCAAGACCCGAAACAACGAAACAACACACACACATGACGTA 5097
QY 410 Tyr-----TyrValAsp-----LeuGluGlnGlnThr 418
DB 5098 TCAACACTATGCTCAAGCGTTTAACTGACTGGGTAACTCACTAGGTGAAGACGCC 5157
QY 419 GluGlyValLeuGlnGlu 424
DB 5158 TCTTCTGACTAACGAA 5175
RESULT 14

```

```

US-08-545-528D-1
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; Patent No. 6537773
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: P0193P1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; PRIOR FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

```

```

Alignment Scores:
Pred. No.: 1,486+03 Length: 580073
Score: 99.00 Matches: 88
Percent Similarity: 38.58% Conservative: 59
Best Local Similarity: 23.10% Mismatches: 158
Query Match: 4.44% Indels: 76
DB: 4 Gaps: 19
US-10-069-062-7 (1-432) x US-08-545-528D-1 (1-580073)

```

```

QY 4 AlaPheSerAlaProGlyLysAlaPheLeuAlaGlyTyrLeuValLeuGluProIle 23
DB 477831 GCTTTAGTACATCAATTAATAATTTCTTACCA-----CTGAAACAA 477875
QY 24 TyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrPro----- 41
DB 477876 ATACCTGTGTTACTTATGTTATCATCAATCAAAATTCAGATCTATCTTAAAACTTA 477935
QY 42 LysGlyThrSerLeuLysGluSerArgIleLysIleSerSerProGlnPheAlaAsnGly 61
DB 477936 AAAAAATCTGCAATGTTTAATGATGATGATGATGATGATGATGATGATGATGATGAT 477992
QY 62 GluTyrGluTyrHisIleSerSerAsnThrGlnLysProArgGluValGlnSerArgIle 81
DB 477993 AGCATTGAAGACAAAGTTTAAAGTTCTCTCAAAACCTCGTCAAC---CAATTTTGTGA 478049
QY 82 AsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnProThrGluAla 101
DB 478050 GAACCTGAATCTTGAAGTCTAGATCTATTTATTTAGAGATTA-----TCAACTTCT 478103
QY 102 PheAspLeuGluIle-----IleIleTyrSerAspProGlyTyrHisSerGlnGlu 118
DB 478104 TTTACACAGAAATTCAAAAAGAAATCATCCAGTTTACCTGTTTCAAAATGACAGA 478163
QY 118 ----- 118
DB 478164 ATTAAAGTTTGTGTTACGCTATTAATGATGATGATGATGATGATGATGATGATGATGAT 478223
QY 119 AspThrGluThrLysThrSerSerAsnGlyLysThrPheLeuTyrHisSerArgIle 138
DB 478224 ACACCTGAAGACAAAGTTTAAATGAAAAC-----TGAATTTTGTGACAA 478268
QY 139 IleThrGluVal-----GluLysThrGlyLeuGlySerSerAlaGlyLeuVal 154
DB 478269 ATTATGACACAGAGGTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATATAT 478328
QY 155 SerValValAlaThrSerLeuLeuSerHisPheIleProAsnValIleSerThrAsnLys 174
DB 478329 GCT-----GCTTTAAATTAATTAATAAA---CCACCATTTATTTGCAACGATATGAG 478379

```

QY 175 AaplleuHiaasnValalaglInleahlsCysTyralaglInlylsylegIser 194
Db 478380 GCTTAT-----ATTGGGGTATGATTAATGATTAGTACTTAAACATCACTGAT 478430
QY 195 GlyPheAspValAlaThr-----AlaIleTyrglyLeu---IleValTyrgArgPhe 211
Db 478431 CCATACCGTTTGTATACATCCAGACGAGATATAGACTATGATGTAAGAAATGACATGTT 478490
QY 212 GlnProAlaLeuIleAsnAspValPheGlnValleuGluSerAspProGluLysPhePro 231
Db 478491 CAAAGACGGCTCTTATTAATAAAAGCTTTGAACTTGTTACAGATTAATAAAACATATGAA 478550
QY 232 ThrGluLeuLysLeuIleGluSerAsnTrpGluGluLysHisGluArgCysThrLeu 251
Db 478551 TTGTCCTTAAAG-----GAAAGAAAAACAGGAATTAAT 478589
QY 252 ProTyrglyIleLysLeuLeuMetGlyAspValTyrglyGlySerGluThrProLysLeu 271
Db 478590 TCATTTTAAAAACACTCACTAGGCAAGTTAAAGCATTAATTCCTATTAAT 478649
QY 272 ValSerArgValleuGlnTrpLysGluLysProGluGluSerSerValValTyrgAsp 291
Db 478650 ACCGCTCAATCACT-----TATGAC 478670
QY 292 GlnLeuAsnSerAlaAsnLeuGlnPheMetCysGluLeuArgLumetArgGluLysTyrg 311
Db 478671 TTCACAAACGAAAGGAATAATTAATTAATGATTAATGATTAATTAATTAATTAATTAAT 478730
QY 312 AspSerAspProGluThrTyrglyIleLysGluLeuAspHisSerValGluProLeuThrVal 331
Db 478731 CAATTAGATTCAAGAAACA---CTTAAACAAATTTGAATTAATTAATTAATTAATTAAT 478787
QY 332 AlaIleLysAsnIleArg-----LysGlyLeuGlnAlaLeuThrGlnLysSerGluVal 349
Db 478788 ATTAAGAAATGAATAAGATTATTTTAAAGGCTTTAAGATAATTAATTAATTAATTAAT 478844
QY 350 Pro 350
Db 478845 CCT 478847

RESULT 15
US-09-214-808-1
; Sequence 1, Application US/09214808A
; Patent No. 6475793
; GENERAL INFORMATION:
; APPLICANT: Rosenthal, Andre
; APPLICANT: Freiberg, Christoph
; APPLICANT: Perret, Xavier Philippe
; APPLICANT: Broughton, William John
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
; Patent No. 6475793
; TITLE OF INVENTION: Plasmid
; FILE REFERENCE: CARP0068
; CURRENT APPLICATION NUMBER: US/09/214,808A
; CURRENT FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: PCT/IB97/00950
; PRIOR FILING DATE: 1997-07-10
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 536165
; TYPE: DNA
; ORGANISM: Rhizobium
US-09-214-808-1

Alignment Scores:
Pred. No.: 1.71e+03
Score: 98.00
Percent Similarity: 35.42%
Best Local Similarity: 20.69%
Query Match: 4.39%
De: 4
Length: 536165
Matches: 66
Conservative: 47
Mismatch: 95
Indels: 111
Gaps: 14

US-10-069-062-7 (1-432) x US-09-214-808-1 (1-536165)
QY 78 GlnSerArgIleAsnProPheLeuGlnAlaThrIlePheIleValleuAlaTyrgIleGln 97
Db 313999 CAATCAGAGATTCTGCTTTTGTAAAGTTCCAGTCTTTTTCGGCGCCAGTCTTCAT 314058
QY 98 ProThrGlnAlaPheAspLeuGluIleIleIleTyrgSerAspProGluTyrgHisSerGln 117
Db 314059 CCCCGGTGTGATGAAAGCTGAACTGGAAGAAATGGCGAGATGCTGTTTGGCCCAAT 314118
QY 118 GluAspThrGluThrLysThrSerSerAsnGlyGluLysThrPheLeuTyrgHisSerArg 137
Db 314119 GTAACTTCCCGCCATCTTCTTCATGGA-----CGG 314154
QY 138 AlaIleThrGlnValGlnLysThrGlyLeuGlySerSerAlaGlyLeuValSerValVal 157
Db 314155 TTTCAGCCGATATGAG 314214
QY 158 AlaThrSerLeuLeuSerHisPheIleProAsnValIleSerThrAsnLysAspIleLeu 177
Db 314215 GCCGCTCAG-----AAA 314226
QY 178 HisAsnValAlaGlnIleAlaHis-----CysTyrgAlaGlnLysLysIleGly 193
Db 314227 AGAATATGCGAGACTCGCTTATGTTGCACTGTGCGGAGGCTCAGACAGCGCGAC 314286
QY 194 SerGlyPheAspVal----- 198
Db 314287 GCCGCGTGACATCATTAATCTCAATCTCGGCTGAAAGCTCGGAGGAGCTGTGCGAAGC 314346
QY 199 -----AlaThrAlaIleTyrgIleLysIleValTyrgArgArgPhe 211
Db 314347 CGCACAGAGCTCAGCTGAGCGACAGCGGCTAGATGGAAGATATTTTCCGTACATA 314406
QY 212 GlnProAlaLeuIleAsnAspValPheGlnValleuGlnSerAspProGluLysPhePro 231
Db 314407 CGG---GCCATCTCGAGAGATATCTTTGTCTCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 314463
QY 232 ThrGluLeuLysLeuIleGluSerAsnTrpGluGluLysHisGluArgCysThrLeu 251
Db 314464 GACCAATG----- 314472
QY 252 ProTyrglyIleLysLeuLeuMet-----GlyAspValTyrglySerGluThr 268
Db 314473 ---TACAGATATCCGACTCTCCAGGACAGACGATACATC---GGCGGCTCC----- 314520
QY 269 ProLysLeuValSerArgValleuGlnTrpLysGluLysProGluGluSerSerVal 288
Db 314521 ---ACCATCGATCGGCTG-----CTCTCGAAGCGCTCGATG 314553
QY 289 ValTyrgAspGlnLeuAsnSerAlaAsnLeuGlnPheMetCysGluLeuArgGluMetArg 308
Db 314554 ---GAGCAGATTACATCGCT-----TTCAATCCGTCGAGCACTTGGCA 314595
QY 309 GluTyrgAspSerAspProGluThrTyrgIleLysGluLeuAspHisSerValGlu--- 327
Db 314596 AAGCGAATGAT-----GAACTGAGCGCCAGCTGAGCAT 314631
QY 328 -----ProLeuThrValAlaIleLys-AsnI 336
Db 314632 GTGCAGCGCGAATATCTGATCGTGTGTCGTCGATCTGATCCGAAATGACAGAACTG 314691
QY 336 eArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAsp 354
Db 314692 ATCGAAAAGCTTGGCGCTTATGTCCTCCGGCTCATGATTAACGAGACAGCGGAA 314746

RESULT 16
US-09-134-001C-1546
; Sequence 1546, Application US/09134001C
; Patent No. 6580370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

```

/ TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: GTC-007
/ CURRENT APPLICATION NUMBER: US/09/134,001C
/ CURRENT FILING DATE: 1998-08-13
/ PRIOR APPLICATION NUMBER: US 60/064,964
/ PRIOR FILING DATE: 1997-11-08
/ PRIOR APPLICATION NUMBER: US 60/055,779
/ PRIOR FILING DATE: 1997-08-14
/ NUMBER OF SEQ ID NOS: 5674
/ SEQ ID NO 1546
/ LENGTH: 1521
/ TYPE: DNA
/ ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1546

Alignment Scores:
Pred. No.: 0.128      Length: 1521
Score: 97.50      Matches: 67
Percent Similarity: 38.64%      Conservative: 35
Best Local Similarity: 25.38%      Mismatches: 99
Query Match: 4.37%      Indels: 63
DB: 4      Gaps: 14

US-10-069-062-7 (1-432) x US-09-134-001C-1546 (1-1521)
QY 43 GlyThSerLeuYsgIuSerAArgIleYsIleSerSerProGlnPheAlaengIyGlu 62
DB 19 GGAAGTCGTTTGAATGCACTTGAATTAATTCATTAATGCAATATGATGAGAA 78
QY 63 TrpGluYrHisIleSerSerAenthrgIuYsProArgIuValGlnSerAArgIleasn 82
DB 79 TGG-----GTTGAAAGTTCAATAAACAACAAGAGATAT-----ATTAAT 120
QY 83 ProPheLeuGluAlaHriIlePheIleValLeuAlaYrIleGlnProThrgIuAlaPhe 102
DB 121 CCTTCAATCAATCAACAACTCTCAGTGA-----GCTGAAGCACTAAAGAA--- 168
QY 103 AspLeuGluIleIleIleYrSer-----AspProGlyYrHisSerGln 117
DB 169 GATGTTGAAAGACATTTTACGCTGAGAAAGATCTTTCGAACAGCGTAATGATCATT 228
QY 118 GluAspThrgIuThrgIuYrHisSerSerAengIyGluYsThrgPheLeuYrHisSerArg 137
DB 229 GAAACAGTCAAGTACGAG-----GCTAAAGAAAGTGT-----AGA 261
QY 138 AlaIleThrgIuValGluYsThrgIyLeuGlySerSerAlaGlyLeuValSerValVal 157
DB 262 GCGGTGCTGATAAATTAAGAAATAGAGAAAGATTAGCTAAATTAAGAACATTAGAC 321
QY 158 AlaThrSerLeuSerHisPheIleProAsnValIleSerThrAsnIysAspIleLeu 177
DB 322 ACTGCTAAACCTTGAAGAACTCTATGCTGATATG-----GATGATAT 366
QY 178 HisAsnValAlaGlnIleAlaHisCysYrAlaGlnIlySylIleGlySerGlyPheAsp 197
DB 367 CATTAATGTTTATGATTTTCTGCTGTTAGCTGATTAAGATGCGGTGAATTAATCAAT 426
QY 198 -----ValAlaThrAla 201
DB 427 TCACCTATTCTTAATCTGAAGTAAAGTAGTTAAAGAACCTGTAGGTGTTTACTCA 486
QY 202 Ile-----YrGlyLeuIle-----ValYrArgArpPheGlnProAlaLeuIle 216
DB 487 ATTACACCTTGAACACTTCATTAATTCGAAGCATCTTGAAGAAATTCGCCACGCTTTAGCA 546
QY 217 AsnAspValPheGlnValLeuGlnSerAspProGlnIysPheProThrgIuLeuYsYs 236
DB 547 ACAGGTGCTCATTAAGTATGAAA-----CCAAGTAAATTAATCCG 588
QY 237 LeuIleGlnSerAenthrgIuYsHisGluArgCysThrgLeuProYrGly---Ile 255
DB 589 TTAAACAACAATTCGTATTGTAATTGATGAGAGAAAGTGTTCCTTAAGAGAACATTT 648

```

```

QY 256 LysLeuLeuMet-----GlyAspValIyGlyGlySerGluThrgPro 269
DB 649 AATTAGTACTGTGGCTGATCAGATCAGAGTGGCGGACGTATGTCAGTCAAGAGATC 708
QY 270 LysLeuValSer 273
DB 709 GATTAGATTCA 720

RESULT 17
US-09-495-406-12
/ Sequence 12, Application US/09495406
/ Patent No. 6503744
/ GENERAL INFORMATION:
/ APPLICANT: Gilbert, Michel
/ APPLICANT: Makarchuk, Warren W.
/ APPLICANT: National Research Council of Canada
/ TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
/ TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
/ FILE REFERENCE: 019633-000110US
/ CURRENT APPLICATION NUMBER: US/09/495,406
/ CURRENT FILING DATE: 2000-01-31
/ PRIOR APPLICATION NUMBER: US 60/118,213
/ PRIOR FILING DATE: 1999-02-01
/ NUMBER OF SEQ ID NOS: 35
/ SOFTWARE: Patencin Ver. 2.1
/ SEQ ID NO 12
/ LENGTH: 1044
/ TYPE: DNA
/ ORGANISM: Campylobacter jejuni
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(1044)
/ OTHER INFORMATION: beta-1,4 N-acetylglactosaminyl (GalNAc)
/ OTHER INFORMATION: transferase from C. jejuni strain OH4364 (ORF 5a)
/ OTHER INFORMATION: of LOS biosynthesis locus)
US-09-495-406-12

Alignment Scores:
Pred. No.: 0.0905      Length: 1044
Score: 96.50      Matches: 65
Percent Similarity: 33.61%      Conservative: 55
Best Local Similarity: 18.21%      Mismatches: 120
Query Match: 4.33%      Indels: 117
DB: 4      Gaps: 13

US-10-069-062-7 (1-432) x US-09-495-406-12 (1-1044)
QY 41 ProYsgIyThrSerLeuYs-----GluSerArgIleYsIleSer 54
DB 127 CCAAAAACCTCTGTCGAATTAATAATCAACAATGAAGATTATTAATTAATTAAT 186
QY 55 SerProGlnPheAlaAsnGlyGluTrpGluYrHisIleSerSerAenthrgIuYsPro 74
DB 187 AAAGCTATTATAGGGAGGGGCGATTAAGAGATTTT-----AATTATGATAAAAATCT 240
QY 75 ArgGluValGlnSerArgIleAsnProPhe----- 84
DB 241 AAAGTCCAAATCTCTTGAATCTTGGGCTTTTATACGTAAGTAAATGAAGTATT 300
QY 85 ---LeuGluAlaThrIlePheIleValLeuAlaYrIleGlnProThrgIuAlaPheAsp 103
DB 301 ACCTTAAGAGCTTCTCTGCAAGACATTAATGCTGCTATCCAA----- 342
QY 104 LeuGluIleIleIleYrSerAsp----- 111
DB 343 AGAGGTATATGATATATATGATATGATCCAGATGAGAGTAAATTAATTCAGAAATT 402
QY 112 -----ProGlyYrHisSerGlnGluAspThrgIuThrgIy 123
DB 403 TCGAAACAATATCCTTCAATTTATACCAATAAATATCTTATGAATTAATTAATCAAAAC 462
QY 124 ThrSerSerAengIyGluYsThrgPheLeuYrHisSerArgAlaIleThrgIuValGlu 143

```

```

Db      463 CCAAAATCAGAGAAAATACTCTATAGCTATTAT----- 498
Qy      144 LysThrGlyLeuGlySerSerAlaGlyLeuValSerValAlaThrSerLeuSer 163
Db      499 -----AATTATGTTGCA 510
Qy      164 HisPheIleProAsnValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIle 183
Db      511 AGTTTATACCA-----AAAGATGAGTGCTTATAAAATATAGATG 552
Qy      184 AlaHisCysThrIleGlnLysLysIleGlySerGlyPheAspValAlaThrAlaIle 203
Db      553 GATCATATCTATATGCTTAAACCTTATATAAAGCTTATATACCAAAAACAAATAT 612
Qy      204 GlyLeuIleValIleValIleValIleValIleValIleValIleValIleValIle 223
Db      613 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 672
Qy      224 GluSerAspProGlnLysPheProThrGlnLeuLysLeuLysLeuLysLeuLysLeu 243
Db      673 AAAGATATAT-----AATGCAATATATTTG 696
Qy      244 GlyLysHisGlnArgCysThrLeuProIleValIleValIleValIleValIleVal 263
Db      697 AAAGAACAGAGAGATGCTTGTATCAATATATATATATATATATATATATATATAT 753
Qy      264 GlyLysSerGlnThrProLysLeuValSerArgVal-----LeuGlnThrLysLysGlnLys 282
Db      754 -----TTAATTGACAGATCAATATATATATATATATATATATATATATATAT 792
Qy      283 ProGlnLysSer----- 287
Db      793 AAACAAAGTTTCTTCAATATATATATATATATATATATATATATATATATATATAT 852
Qy      288 ValValIleValIleValIleValIleValIleValIleValIleValIleValIle 307
Db      853 TTAATTGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 912
Qy      308 Arg-----GlnLysPhe-----AspSerAspProGlnLysThrIleLysGln 321
Db      913 GATATTATATATATATATATATATATATATATATATATATATATATATATATATATAT 972
Qy      322 LeuAspHisSerValGlnProLeuThrValAlaIleLysAsnIleArgLys 338
Db      973 ATTAATCATATAAATAGAACCTTCTATGATTTCAAAAGAACTCTATAAAAAA 1023

RESULT 18
US-08-347-563A-1
; Sequence 1, Application US/08347563A
; Patent No. 5935810
; GENERAL INFORMATION:
; APPLICANT: THE ROCKEFELLER UNIVERSITY
; TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING NUCLEIC
; TITLE OF INVENTION: ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/347,563A
; FILING DATE: No. 5935810eember 30, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/292,345

```

```

; FILING DATE: August 17, 1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Bsq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-087 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2793 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; DESCRIPTION: Murine ob cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Murine
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 57..560
; US-08-347-563A-1

Alignment Scores:
Pred. No.: 0.457 Length: 2793
Score: 96.50 Matches: 87
Percent Similarity: 35.43% Conservative: 54
Best Local Similarity: 21.86% Mismatches: 157
Query Match: 4.33% Indels: 101
DB: Gaps: 13

US-10-069-062-7 (1-432) x US-08-347-563A-1 (1-2793)
Qy      68 SerSerAsnThrGlnLysProArgGlnValGlnSerArgIleAsnProPheLeuGlnAla 87
Db      11 TCCAGCAGCTGCAAGGTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 70
Qy      88 ThrIlePheIleValIleValIleValIleValIleValIleValIleValIleValIle 107
Db      71 CCGTGGCGGTCCTGCGGCTTGTGTC-----CTATCTGTCTTA 109
Qy      108 IleTyrSerAsp-ProGlyTyrHisSerGlnLysPheThrGluThr-----LysThr-S 125
Db      110 TGTTCACAGCAGTGCCTATCCAGAAAGTCAGAGATGACACAAACCTTCATCAAGACCAT 169
Qy      125 erSerAsnGlyGlnLysThrPheLeuTyrHisSerArgAlaIleThrGlnValGlnLys 145
Db      170 TGTTCACAGCAGTGCCTATCCAGAAAGTCAGAGATGACACAAACCTTCATCAAGACCAT 228
Qy      145 hrgLysLeuGlySerSerAlaGlyLeuValSerVal----- 156
Db      229 CTGGTTCGATTCATTCCTGCGCTTCAACCCATTCGAGTTTGCCAAAGATGACACGA 288
Qy      157 -----ValAlaThrSerLeuSerHis-----PheIleP 167
Db      289 CTGTGACAGTATATCAACAGGCTCTCAACAGCCTGCTCCCAAAATGCTGCAGATAG 348
Qy      167 roAsnValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCys 187
Db      349 CCAATGACCTGAGATATTCAGAGACCTCTCCATCTGCTGCGCTTCTCCAAAGCTCT 408
Qy      187 ValAlaGlnLysLysIleGly-----SerGlyPheAsp-----ValAlaThr 201
Db      409 CCGTGCCTCAGACAGAGGCTGCGCAAGACCAAGAGAGCTGAGAGGCTCTCGAAGCT 468
Qy      201 lalleTyrGly-----LeuIleValIleValIleValIleValIleValIleValIle 219
Db      469 CACTCTACTCCACAGAGGTGTGCTTGTAGACAGGCTGCAAGGCTCTCTG---CAGGACA 525

```



```

Db      409  CCCTGCTCAGACCACTGCTGCTGAGAGCCGCTGAGCTGCTGGAAGCT 468
      201  lalleTyrgly-----LeuileValTyraGrphglnProalaleuileAsnAspV 219
      469  CACTTACTCCACAGAGGTGGTGGCTTTGAGCAGGCTCTCTG--CAGAGCA 525
      219  aPheGlnValleugJuserAspProglu-----L 229
      526  TTCTTCAACAGTTGATGTTAGCCCTGAGTTCGAACTTCAAGGCCACAGGCTCCCA 585
      229  yPheProThrGlnleuLysLys-LeuileGlnSerAsnTrpGlnLysLysLysLys 248
      586  AGATCATGTAGAGGAGAAACCTTGCTTCCAGGGGCTTCCAGAGAGAGAGAGCAG 645
      249  CysThrLeuProTyrglyLysLysLeuLeuMerGlyAspVallyLysLysLysLys 268
      646  TGCACACATCCATCATTCATTCCTCTCTCTGAGACCAACCCATCCAAAGGCATGACT 705
      269  ProLysLeuValSerArgValleugIntTplyLysGlnLysProgluLysSerVal 288
      706  CCACATGCTTGAAGTAACTTATCCACACACTTCAAGACAGAGAGAGGCGCAGC-- 762
      289  ValTyraSPGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGlnLeuArgLysLys 308
      762  ----- 762
      309  GlnLysTyraSPSerAspProgluThrTyrlleLysGlnLeuAspHisSerValGlnPro 328
      763  -----CTGACAGAGGAGGACTCTCAGCT 783
      329  LeuThrValAlaIleLysAsnIleArgLysGlnLeuGlnAlaLeuThrGlnLysSerGln 348
      784  AGTTCCTTCAGCAGATAGATAGATAGAGCCATCCCATCCCTCATGCTCCAGCTCCCG 843
      349  ValProLleGlnProAspValGlnThrGlnLeuLeuAspArgCysGlnLysLysLys 368
      844  GTACATGTTCTCCGCGGTACACGCTTCTGCGGCGCCAGAGAGAGTGAAGTA----- 897
      369  CysValGlyGlyValValProGlyAlaGlyGlyTyraSPAlaIleAlaValleuValLeu 388
      898  -----GGATGGGTAGAGCCTTGGCTGTCTCAGAGCTTTGG 936
      389  GlnAsnGlnValGlyAsnPhe-----LysGlnLysThrLeuGln 401
      937  GAGCACCGGAGAGGCTGATCCACACACAGCTGGAACCTCCCAAGACAGACACAGTGA 996
      402  AsnProAspTyraSP---PheHisAsnValTyraSPValAspLeuGln 416
      997  GCACCTATTATTATTATTCGACATTCATTTGGATGGATGATGAGCA 1044

```

RESULT 20
 US-08-488-214A-1
 ; Sequence 1, Application US/08488214A
 ; Patent No. 6124439
 ; GENERAL INFORMATION:
 ; APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENZA,
 ; APPLICANT: MARGHERITA MARPEI, JEFFREY HALAS, KETAN GAJWALA, AND STEPHEN K. BURL
 ; TITLE OF INVENTION: OB POLYPEPTIDE ANTIBODIES AND METHOD OF MAKING
 ; NUMBER OF SEQUENCES: 99
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Klauber & Jackson
 ; STREET: 411 Hackensack Avenue
 ; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07601
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,214A
FILING DATE: JUNE 7, 1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/438,431
FILING DATE: May 10, 1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/347,563
FILING DATE: No. 612443September 30, 1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/292,345
FILING DATE: August 17, 1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-087 CIP 2D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2793 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
DESCRIPTION: Murine ob cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Murine
FEATURE:
NAME/KEY: CDS
LOCATION: 57..560
US-08-488-214A-1
Alignment Scores:
Pred. No.: 0.457 Length: 2793
Score: 96.50 Matches: 87
Percent Similarity: 35.43% Conservative: 54
Best Local Similarity: 21.86% Mismatches: 157
Query Match: 4.33% Indels: 101
DB: 3 Gaps: 13
US-10-069-062-7 (1-432) x US-08-488-214A-1 (1-2793)

```

QY 68 SerSerAsnThrGlnLysProArgGlnValGlnSerArgIleAsnProPheLeuGlnAla 87
 DB 11 TCCAGCAGCTGACAGGTGCAAGAGAGAGATCCACGAGAGAGAAATGTGCTGAGACC 70
 QY 88 ThrIlePheIleValleuAlaTyrlleGlnProThrGlnAlaPheAspLeuLysLys 107
 DB 71 CCTGTGCGGTCTCTGAGCTTGGTC-----CTATCTGTCTTA 109
 QY 108 IleTyraSP---ProGlyTyraSPHisSerGlnLysAspThrGlnThr-----LysThr-S 125
 DB 110 TGTTCACACAGTGGCTTATCCAGAGAAAGTCCAGAGTACACCAAAACCTCATCAAGACCAT 169
 QY 125 erSerAsnGlyGlnLysThrPheLeuTyraSPHisSerArgAlaIleThrGlnValGlnLys 145
 DB 170 TGTACCCAGATCAATGATTCATTCACACCGCAGCGGTA--TCCGCCAAGCAGAGGCTCA 228
 QY 145 hrgLysLeuGlySerSerAlaGlyLeuValSerVal----- 156
 DB 229 CTGGCTTGACTTCAATTCCTGGGCTTCAACCCCATTCGAGTTTGTCCAAAGTGCACCA 288
 QY 157 -----ValAlaThrSerLeuSerHis-----PheIleP 167

```

Db      CTCTGACAGTCTATCAACAGAGTCTCCACAGAGCTGCTCCCAAAATGCTGCAGATAG 348
Qy      167 roasnVal1leserThrAsnLysAspPheLeuHisAsnValAlaGlnIleAlaHisCysT 187
Db      349 CCATACCTGAGGAAATCTCCGAGACCTCTCCATCTGCTGCGCTTCTCCAGAGCTGCT 408
Qy      187 yralaGlnLysIleGly-----serGlyPheAsp-----ValAlaThrA 201
Db      409 CCTGCTCTGACAGACAGTGGCTCTCCAGAGCCAGAGCCCTGAGTGGCTCTGGAAGCCT 468
Qy      201 lalleTyrgly-----LeuIleValTyraArgPheGlnProAlaLeuIleAsnAsp 219
Db      469 CACTCTACTCTCCACAGAGGTGTGCTTTGAGCAGGCTGCGAGGCTCTCTG---CAGAGA 525
Qy      219 a1PheGlnValLeuGlnSerAspProGlu-----L 229
Db      526 TTCTTCAACAGTTGAGTGTAGCCCTGAAAGCTGAAAGTTCAAAAGCCACAGGCTCCCA 585
Qy      229 ysfPheProThrGlnLeuLysLys-LeuIleGlnSerAsnTrpLysGlnLysHisGlnArg 248
Db      586 AGAATCATGTAGAGGAGAAACCTTGGCTTCCAGGGGTCTTCAGAGAGAAAGAGAGCCATG 645
Qy      249 CysThrLeuProTyrglyIleLysLeuLeuMetGlyAspValLysGlySerGlnThr 268
Db      646 TGCACACATCATCATTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
Qy      269 ProLysLeuValSerArgValLeuGlnTrpLysGlnLysProGlnGlnSerVal 288
Db      706 CCACAAATGCTTGACTAAGTTATCCACAACTTCATGAGACACAGAGAGGGGCTCAC-- 762
Qy      289 ValTyraSpGlnLeuAsnSerAlaAsnLeuGlnPheMetLysLeuArgGlnMetArg 308
Db      762 ----- 762
Qy      309 GlnLysTyraSpSerAspProGlnThrTyrlLysGlnLeuAspHisSerValGlnPro 328
Db      763 -----CTGCAGAGGGGACCTCTCACCT 783
Qy      329 LeuThrValAlaIleLysAsnIleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGln 348
Db      784 AGTTCTTCACAGATGAGATGAAGAGCCATCCATCCCTCCATGTCACCTGCTCCGG 843
Qy      349 ValProIleGlnProAspValGlnThrGlnLeuLeuAspArgCysGlnGlnIleProGly 368
Db      844 GTACATGTTCTCCCGGAGGACACGCTGCTGCGGCCAGAGAGGTAGAGTA----- 897
Qy      369 CysValGlyGlyValValProGlyAlaGlyLysTyraSpAlaIleAlaValLeuValLeu 388
Db      898 -----GGGATGGGTAGAGCCTTTGGCTGTCTCTCAGAGTCTTTGG 936
Qy      389 GlnAsnGlnValGlyAsnPhe-----LysGlnLysThrLeuGln 401
Db      937 GAGCAGCCGTGAGAGCTGCATCCACACAGCTGGAATCTCCAGAGCAGCAGAGTGA 996
Qy      402 AsnProAspTyra---PheHisAsnValTyrlTyrlValAspLeuGln 416
Db      997 GCATCTATTATTATTCTGCAATCTTATTTGATGATGATGAGCA 1044

```

```

RESULT 21
US-08-488-208A-1
; Sequence 1, Application US/08488208A
; Patent No. 612448
; GENERAL INFORMATION:
; APPLICANT: THE ROCKEFELLER UNIVERSITY
; TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauder & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey

```

```

; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,208A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/485,943
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: 08/438,431
; FILING DATE: May 10, 1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/347,563
; FILING DATE: No. 612448ember 30, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/292,345
; FILING DATE: August 17, 1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-087 CIP21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2793 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; DESCRIPTION: Murine OB cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Murine
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 57..560
; US-08-488-208A-1
;
; Alignment Scores:
; Pred. No.: 0.457 Length: 2793
; Score: 96.50 Matches: 87
; Percent Similarity: 35.43% Conservative: 54
; Best Local Similarity: 21.86% Mismatches: 157
; Query Match: 4.33% Indels: 101
; DB: 3 Gaps: 13
;
; US-10-069-062-7 (1-432) x US-08-488-208A-1 (1-2793)
;
; Qy      68 SerSerAsnThrGlnLysProArgGlnValGlnSerArgIleAsnProPheLeuGlnAla 87
; Db      11 TCCAGAGCTGCAAGTGCAGAGAAGAGATCCACAGAGAGAAATGCTGGAGAC 70
; Qy      88 ThrIlePheIleValLeuAlaTyrlleGlnProThrGlnAlaPheAspLeuGlnIle 107
; Db      71 CCTGTCTCGTCTCTGCTGCTTGTGTC-----CTATCTGTCTTA 109
; Qy      108 IleTyraSp---ProGlyTyrlHisSerGlnLysAspThrGlnThr-----LysThr-S 125
; Db      110 TGTTCAGAGAGTCCATATCCAGAAAGTCCAGAGTACACAAACCTCATCAAGCCAT 169
; Qy      125 erSerAsnGlyGlnLysThrPheLeuTyrlHisSerArgAlaIleThrGlnValGlnLysT 145

```



```

Db      170 TGCACACAGATCAATGACATTTCACACAGCAGTGGTA-TCGCCACAGAGGGTCA 228
      145 hcglyleuylserialaglyleuValSerVal----- 156
      229 CTGGCTTGACCTTCTCTGCTTCCACCCCTTGAGTTGTCMAAGATGACACAGA 288
      157 -----ValAlaThrSerleuSerHis-----PheIleP 167
      289 CTCTGGCACTATCAACAGCTCTCAACAGCTGCTTCCAAAATGCTGCAGATNG 348
      167 roasnValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysT 187
      349 CCAATGACCTGGAGAAATCTCCAGACCTCTCCATCTGCTGCTTCCAGAGCTGCT 408
      187 yAlaGlnLysLysIleGly-----SerGlyPheAsp-----ValAlaThrA 201
      409 CCCTGCTCAACACAGTGGCTGACAGAGCCAGAGACCTGATGGCTGCGAAGCCT 468
      201 laIleTyGly-----LeuIleValTyArgArgPheGlnProAlaLeuIleAsnAspV 219
      469 CACTCTACTCCACAGAGGTGGTGGCTTTGAGCAGCTGCGAGGGCTCTG---CAGACA 525
      219 aIleGlnValleuGlnSerAspProGly-----L 229
      526 TTCTTCAACAGTTGATGTTAGCCCTGAGATGCTGAAGTTCAAGGCCACAGGCTCCA 585
      229 yPheProThrGlnLeuLysLys-LeuIleGlnSerAsnTrpGlnLysIleGlnA 248
      586 AAATCATGATGAAGGAGAAACCTTGCTTCCAGGGGCTTCCAGGAAGAAGAGCCAG 645
      249 CysThrLeuProTyGlyIleLysLeuLeuMetGlyAspValIleGlySerGluThr 268
      646 TGCACACATCCATCATCTTCTCTCTCCCTGCTGACACCCATCCAAAGCATGCT 705
      269 ProLysLeuValSerArgValleuGlnTrpLysGlnLysProGlnLysSerVal 288
      706 CCACATGCTTGAATCAAGTTATCCACACACTTATGAGCACAAGAGAGGGCCAGC--- 762
      289 ValTyAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGlnLeuArgLysLeu 308
      762 ----- 762
      309 GlnLysTyAspSerAspProGluThrTyIleLysGlnLeuAspHisSerValGluPro 328
      763 -----CTGCAGAGGGGACTCTCACT 783
      329 LeuThrValAlaIleLysAsnIleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGln 348
      784 AGTTCCTTACGAAAGTGAATGAAGACCATCCCATCCCTCATGTCACACCTGCTCGG 843
      349 ValProIleGlnProAspValGlnThrGlnLeuLeuAspArgCysGlnIleProGly 368
      844 GTACATGTTCTCTCCGGGTACACGCTTGGCTGGCCGCCAGAGAGCTGAGTA----- 897
      369 CysValGlyLysValValProGlyAlaGlyLysTyAspAlaIleAlaValleuValleu 388
      898 -----GGGATGGGTAGAGCTTTGGCTGTGCTGAGGCTCTTGG 936
      389 GluAsnGlnValGlyAsnPhe-----LysGlnLysThrLeuGln 401
      937 GAGCACCGTGAAGGCTGCATCCACACACAGCTGAAGAACTCCAGAGAGACAGCATGGA 996
      402 AsnProAspTy---PheHisAsnValTyTrpValAspLeuGlnGlu 416
      997 GCACCTATTATTATTCTGCACTCTATTGATGATGATGATGAAACAA 1044

```

RESULT 22

US-08-483-211A-1

Sequence 1, Application US/08483211A

Patent No. 6309853

GENERAL INFORMATION:

APPLICANT: THE ROCKEFELLER UNIVERSITY

```

      TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING
      TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC
      TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC
      NUMBER OF SEQUENCES: 98
      CORRESPONDENCE ADDRESS:
      ADDRESS: Klauber & Jackson
      CITY: Hackensack
      STATE: New Jersey
      COUNTRY: USA
      ZIP: 07601
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      OPERATING SYSTEM: IBM PC compatible
      SOFTWARE: Patent Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/483,211A
      FILING DATE: 07-JUN-1995
      CLASSIFICATION: 514
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 08/485,943
      FILING DATE: June 7, 1995
      APPLICATION NUMBER: 08/438,431
      FILING DATE: May 10, 1995
      CLASSIFICATION: 514
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 08/347,563
      FILING DATE: No. 6309853ember 30, 1994
      CLASSIFICATION: 514
      APPLICATION DATA:
      APPLICATION NUMBER: 08/292,345
      FILING DATE: August 17, 1994
      CLASSIFICATION: 514
      ATTORNEY/AGENT INFORMATION:
      NAME: Jackson Esq., David A.
      REGISTRATION NUMBER: 26,742
      REFERENCE/DOCKET NUMBER: 600-1-087 CIP21
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 201 487-5800
      TELEFAX: 201 343-1684
      TELEX: 133521
      INFORMATION FOR SEQ ID NO: 1:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 2793 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
      MOLECULE TYPE: DNA (genomic)
      DESCRIPTION: Murine ob cDNA
      HYPOTHEICAL: NO
      ANTI-SENSE: NO
      ORIGINAL SOURCE:
      ORGANISM: Murine
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 57..560
      US-08-483-211A-1

```

Alignment Scores:

```

Pred. No.: 0.457 Length: 2793
Score: 96.50 Matches: 87
Percent Similarity: 35.43% Conservative: 54
Best Local Similarity: 21.86% Mismatches: 157
Query Match: 4.33% Indels: 101
DB: 4 Gaps: 13

```

US-10-069-062-7 (1-432) x US-08-483-211A-1 (1-2793)

```

      68 SerSerLeuThrGlnLysProArgGlnValGlnSerArgIleAsnProPheLeuGlnAla 87
      11 TCACAGAGCTGCAAGGTGCAGAGAGAGAGATCCAGGGGAGAGAAATGCTGAGAGCC 70

```

88 ThrillePheIleValleuIaTyrlleGlnIleProThrGlnAlaPheAspLeuGluIle 107
 71 CCTGTGTGGCTTCCTGTGGCTTTGGTC-----CTATGTCTTA 109
 108 IleTySerAsp-ProGlyTyrlHisSerGlnIaPheThrGluThr-----LysThr-s 125
 110 TGTTCAGACAGAGCGCTTATCCAGAAAGTCCAGAGTGAACACCAAAACCTCATCAAGACAT 169
 125 erSerAsnGlyGluTyrlThrPheLeuTyrlHisSerIArgAlaIleThrGluValGluYst 145
 170 TGTCCACAGAGATCAATGACATTTCACACAGCGAGTGGTA-TCCGCCAAGCAGAGGATCA 228
 145 hngIleuclYserSerAlaGlyLeuValSerAl----- 156
 229 CTGGCTTGGACCTTCATTCCTGGGCTTCAACCCCATCTGAGTTTGTCCAGATGGACCA 288
 157 -----ValAlaThrSerLeuLeuSerHis-----PheIleP 167
 289 CTCTGGACAGTCTATCAACAGAGCTTCACACAGCGCTTCCTCCAAATATGTCTCAGATTAG 348
 167 roAsnValIleSerThrAsnGlyAspIleLeuHisAsnValAlaGlnIleAlaHisCysT 187
 349 CCAATGACCTGGAGAAATCTCCAGAGACTCTCCATCTGTGGCTTCTCCAGACGTGCT 408
 187 yAlaGlnIleYsrlleGly-----SerGlyPheAsp-----ValAlaThrA 201
 409 CCTGTGCTTCAGACCAAGTGGCTTCGACAGAACCGAGAGCCTGGATGGCGTCTTGAAGCCT 468
 201 IaIleTyrlGly-----LeuIleValTyrlArgIArgPheGlnProAlaLeuIleAsnAspY 219
 469 CACTCTACTCCACAGAGGTGTGGCTTTGAGCAGCGTGCAGGGCTCTGTG--CAGAGCA 525
 219 aIleGlnValleuGluSerAspProGlu-----L 229
 526 TTTCTTCACAGTGTGATGTGTAAGCCCTGGAATGTGAATTCCAAAGGCCACAGCTCCCA 585
 229 yAPheProThrGlnleuLeuYsrls-LeuIleGluSerAsnTrpGluGlnYsrlsGluArg 248
 586 ACAATCATCATTAAGGGAGAGAACTTGGCTTCAGGGGGCTTCAAGAGAAAGAGACCATG 645
 249 CysThrLeuProTyrlGlyIleYsrlsLeuLeuMetGlyAspValYsrlGlySerGluThr 268
 646 TCCACACATTCATTCATTTCTCTCCCTCCGTATGACACCAACCATCCAAAGGCATGACT 705
 269 ProYsrlsLeuValSerArgValleuGlnTrpYsrlYsrlGluYsrlProGluGluSerVal 288
 706 CCACATGCTGTGACTCAAGTATCCACACAACTTCATGAGACAGAGGGGGCAGC-- 762
 289 ValTyrlAspGlnleuAsnSerAlaAsnLeuGlnPheMetYsrlGluLeuYsrlGluMetArg 308
 762 ----- 762
 309 GlnYsrlTyrlAspSerAspProGluThrTyrlleYsrlleuAsnHisSerValGluPro 328
 763 -----CTGCAGAGGGGACTCTACCT 783
 329 LeuThrValAlaIleYsrlsAsnIleArgYsrlYsrlleuGlnAlaLeuThrGlnYsrlsSerGln 348
 784 AGTCTTTCAGCAAGTGAAGATTAAGACCATCCCATCCCTCCATGTCCCACTGCTCCGG 843
 349 ValProIleGlnProAspValGlnThrGlnleuLeuAspArgYsrlGlnIleProGly 368
 844 GTACATGTTCCTCCGGGTGACACGCTTGCTGGCGCCAGGAGAGGTGAGTA----- 897
 369 CysValGlyGlyValValProGlyAlaGlyGlyTyrlAspAlaIleAlaValleuValleu 388
 898 -----GGAGTGGGTAGAGCCTTGGGCTGTCTCAGATCTTTGG 936
 389 GluAsnGlnValYsrlAsnPhe-----LysGlnYsrlThrLeuGln 401
 937 GAGCACCCGTGAAGGCTGCATCCACACAGCTGAAATCCCAAGCAGCAGACAGATGAA 996
 402 AsnProAspTyrl---PheHisAsnValTyrlTrpValAspLeuGluGln 416

```

Db      11 TCACAGAGCTGCAGAGTGCAGAGAGAGATCCAGGAGGAGAAATGCTGAGACC 70
Qy      88 ThrilePheileValleuValleuTyrlleGlnProthrgluVallePheaspLeuGluile 107
      71 CCGTGTGGTCTCTGCTGGCTTGTGC -----CTATGTGCTTA 109
Qy      108 lletYSerAsp-ProGlylYrhisSerGlnlAspThrgluThr-----LysThr-S 125
      110 TGTTCACAGAGAGTCCCTATCCAGAAAGTCCAGATGACACCAAAACCTCATCAAGACCAT 169
Qy      125 eSerAsnGlyGlyuThrPheLeuThrhisSerTrgAlaileThrgluValleGlyuYF 145
      170 TGTTCACAGAGATCAATGACATTCACACAGCGGTGGTA-TCGCGACAGAGAGGTCA 228
Qy      145 hrcGlyLeuGlySerSerAlaGlyLeuValSerAl----- 156
      229 CTGGCTTGACCTTCATTCCTGGGCTTCACCCCATTCCTGAGTTGTTCAGATGACACCA 288
Qy      157 -----ValAlaThrSerleuSerhis-----PheileP 167
      289 CTCTGCGAGCTATCAACAGAGTCTCCACAGCGCTGCTCCCAAAATGTGTGACAGATG 348
Qy      167 roAsnValileSerThraenlyAspPileuThraenValAlaGlnleAlaHisCysT 187
      349 CCAATGACCTGGAGAAATCTCCAGAGACCTCTCATCTGTGGCTTCTCCAGAGAGCTGCT 408
Qy      187 yAlaGlnlySerlyleGly-----SerGlyPheAsp-----ValAlaThrA 201
      409 CCTGCTCTACAGACAGTGGCTGCGAGAGCCAGAGCGCTGAGTGGCTGCGAAAGCTT 468
Qy      201 lailleTyrlly-----LeuileValYrhisArgPheGlnProAlaLeuileAsnAspY 219
      469 CACTCTACTCCACAGAGGTGTGTGGCTTTGACAGAGCGTGCAGAGGCTCTG-----CAGAGA 525
Qy      219 aPheGlnValleuGlnSerAspProGlu-----L 229
      526 TTCTTCACAGTTGATGTAGTTCAGCCCTGAAATGCTGAGATTCAAGGCCACAGGCTCCA 585
Qy      229 yPheProthrgluLeuLysLys-LeuileGlnSerAsnTrgGlnlulysHisGluArg 248
      586 ABAATCATATAGAGGAGAAACCTTGGCTTCAGAGGAGTCTTCAGAGAGAGAGAGCCATG 645
Qy      249 CysThraenProTyrllyleLysLeuLysLysLysAspValleGlyYserGlnThr 268
      646 TGCACACATCCATTCATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
Qy      269 ProLysLeuValSerArgValleuGlnTrpLysLysGluLysProGlnlulysSerVal 288
      706 CCACATGCTTGTACTCAAGTTATCCACACAACTTCATGAGCACAAAGAGGCGCCAGC-- 762
Qy      289 ValTYrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGlnLeuArgGlnMetArg 308
      762 ----- 762
Qy      309 GluLysTYrAspSerAspProGluThrTYrlleLysGlnLeuAspHisSerValGluPro 328
      763 -----CTGCAGAGGAGACTTCACCT 783
Qy      329 LeuThraValAlaileLysAsnleLysGlyLysGlnAlaLeuThrglnLysSerGlu 348
      784 AGTCTTCACAGAGATGAGATGAGCCATCCATCCCTCTCATGCTCCACCTGCTCCGG 843
Qy      349 ValProileGluProAspValGlnThrglnLeuLysAspArgCysGlnGluileProGly 368
      844 GTRKATGTTCTCTCGGTGATACAGCTTGTGCGGCCCGGAGAGAGGTAGGTA----- 897
Qy      369 CysValleGlyValleValProGlyAlaGlylYrAspAlaileAlaValleuValleu 388
      898 -----GGATGGGTAGAGCTTGGCTGTCTGACAGGTCTTTGG 936
Qy      389 GluAsnGlnValleGlyAsnPhe-----LysGlnLysThrglnGlu 401
      937 GAGCACCGTGAAGCTGCATCCACACAGAGCTGAAATCTCCAAAGCAGACACGATGAA 996

```

```

Qy      402 AsnProAspTYr-----PheHisAsnValTYrTrpValleAspLeuGlu 416
Db      997 GCACCTATTATTATTATTCGACATTCATTTTGGATGATGATCAACCA 1044

RESULT 24
US-08-438-431A-1
Sequence 1, Application US/08438431A
Patent No. 6429290
GENERAL INFORMATION:
APPLICANT: JEFFREY M. FRIEDMAN, YIYING ZHANG, RICARDO PROENCA, MARGHERITA MAFFEI,
TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING NUCLEIC ACIDS AND PR
NUMBER OF SEQUENCES: 99
CORRESPONDENCE ADDRESS:
ADDRESS: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/438,431A
FILING DATE: May 10, 1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/347,563
FILING DATE: August 17, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/292,345
FILING DATE: August 17, 1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-087 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2793 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
DESCRIPTION: Murine ob cDNA
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Murine
FEATURE:
NAME/KEY: CDS
LOCATION: 57..560
US-08-438-431A-1

Alignment Scores:
Pred. No.: 0.457 Length: 2793
Score: 96.50 Matches: 87
Percent Similarity: 35.43% Conservative: 54
Best Local Similarity: 21.86% Mismatches: 157
Query Match: 4.33% Indels: 101
DB: 4 Gaps: 13

US-10-069-062-7 (1-432) x US-08-438-431A-1 (1-2793)
Qy      68 SerSerAsnThrgluLysProArgGluValGlnSerArgIleAsnProPheLeuGluAla 87

```

```

Db      11 TCACAGAGCTGCAAGAGTGCAGAGAAAGATCCACAGAGGAGAAATGCTGAGACC 70
Qy      88 ThrilehellelealeumalaTyrlleglnProthrhglalaphespleugluille 107
Db      71 CCTGTGCGTTCCTGCTTGGCTTTGGTC-----CTATCTGCTTA 109
Qy      108 lletySerasp-ProglYTyrrhisserglngluAepthrhgluThr-----lyethr-S 125
Db      110 TGTTCAAGCAGTCCCTTCATCCAGAAATCCAGATGACCAAAACCTCATCAAGACAT 169
Qy      125 erseranglYgluYsthrpHeleuTyrrhisserargalalethrhgluValgluYst 145
Db      170 TGCACACAGGATCAATGACATTTCAACAACGACGCTGGTA--TCGCGCAAGCAGAGGCTCA 228
Qy      145 hnglyleuYglYserSerAlaglyleuValserAl----- 156
Db      229 CTGGCTTGGACTTCATTCTCTGGCTTCACCCCATTCGATTCGTCAMAGATGACACGA 288
Qy      157 -----ValAlathrSerleuSerHis-----Pheliep 167
Db      289 CTCTGCAAGCTATCAACAGATCTCTCAACGCTGCTTCCAAAATGCTGACATAG 348
Qy      167 rohanValilleserThrasnlyAspilleuHisasnValAlaglnllealHisCysT 187
Db      349 CCAATGACCTGAGAAATCTCCGAGACCTCTCATCTGCTGCGCTTCCCAAGAGCTGCT 408
Qy      187 yrlaaglnYglYslyllegly-----serglYpHeasp-----ValAlathrA 201
Db      409 CCTGCTCCTACAGACAGTGCCTGACAGAAAGCAGACAGCTGAGCGCTGGAAGCCT 468
Qy      201 lalletYrgly-----LeuileValTyrrhArgpHeaglnProAlaleuileasnAspy 219
Db      469 CACTTACTTCCACAGAGGTGGTGGCTTTGAGCAGAGGCTGACAGGCTGCTG---CAGAGA 525
Qy      219 alPheglnValleuYglYserAspProglu-----L 229
Db      526 TTCCTCAACAGTGGATGTTAGCTGAAAGCTGAAAGCTTCAAAAGCCACACAGCTCCCA 585
Qy      229 yspHeProthrhgluYglYslyleuileglYserAsnTrpYgluYshlgluYarg 248
Db      586 AGAATCATGTAGAGGAGAAACCTTGGCTTCCAGGGGCTTTCAGAGAAAGAGAGCCAG 645
Qy      249 CysThrlaupProTyrglylleYsleuLeuMetGlyAspVallysglyYsergluThr 268
Db      646 TGCACACATCATTCATTCATTTCTCTCCCTCGTAGACACACCCATCCAAAGCAGTACT 705
Qy      269 ProYleuValserArgValleuYgluYsthrpYslysglyYsProgluYglYserSerVal 288
Db      706 CCACAATGCTTGACTCAAGTTATCCACACAACATTTCATGACACAGAGAGGCGCCAGC--- 762
Qy      289 ValTyrraspGlneumenserAlaasnleuYgluYpHeMetYslyleuYgluYmetArg 308
Db      762 ----- 762
Qy      309 GluYsTyrraspSerAspProgluThrTyrlleYsgluYleuYpHisSerValgluPro 328
Db      763 -----CTGACAGAGGAGACTCTCACCT 783
Qy      329 LeuThValAlalleYsasnillearglysglyleuYgluYleuThrhgluYsSerGlu 348
Db      784 AGTTCTTTCAGCAAGTGAAGATAGAGCCATCCATCCCTCCATGTCCTCCACCTGCTCCG 843
Qy      349 ValProillegluProaspValgluYsthrpHeleuYslysglyYsGlnYgluYleProglY 368
Db      844 GTACATGTTCTCTCGGTGGTACAGCTTCCGTGCGGCCCGCCAGAGAGTGA----- 897
Qy      369 CysValglYglYValValProglYAlaglyYTyrraspAlallealeValleuValleu 388
Db      898 -----GGATGCGGTAGAGCCCTTGGGCTGTCTGAGATCTTTGG 936
Qy      389 GluasnGlnValYglYasnPhe-----lysGlnYsthrleuYglu 401

```

```

Db      937 GAGCACCGTGAAGGCTGATCCACACAGACTGGAACCTCCCAACAGCAGATGGA 996
Qy      402 AasnProaspTyrr-----PheHisasnValTyrrpValaspLeuYglu 416
Db      997 GCACCTATTATTATTATTCGACTTCATATTGATGATGATCAAGCA 1044

RESULT 25
US-08-488-225A-1
; Sequence 1, Application US/08488225A
; Patent No. 6471956
; GENERAL INFORMATION:
; APPLICANT: THE ROCKEFELLER UNIVERSITY
; TITLE OF INVENTION: MODULATORS OF BODY WEIGHT, CORRESPONDING
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS, AND DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauder & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,225A
; FILING DATE: June 7, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/483,211
; FILING DATE: June 7, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/438,431
; FILING DATE: May 10, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/347,563
; FILING DATE: No. 6471956ember 30, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/292,345
; FILING DATE: August 17, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-087 CIP2J
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2793 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; DESCRIPTION: Murine ob cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Murine
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 57..560
; US-08-488-225A-1

Alignment Scores:

```

Pred. No.: 0.457 Length: 2793
 Score: 96.50 Matches: 87
 Percent Similarity: 35.43% Conservative: 54
 Best Local Similarity: 21.86% Mismatches: 157
 Query Match: 4.33% Indels: 101
 DB: 4 Gaps: 13

US-10-069-062-7 (1-432) x US-08-488-225A-1 (1-2793)

QY 68 SerSerAsnThrGluLysProArgGluValGlnSerArgIleAsnProPheLeuGluAla 87
 DB 11 TCACAGAGCTGCAGAGTGCAGAGAGAGAGAGATCCACGAGGAGAAATGCTGCGAGACC 70
 QY 88 ThrIlePheIleValLeuAlaTyrIleGlnProThrGluAlaPheAspLeuGluIle 107
 DB 71 CCGTGTGCGTTCCTGCTGCTTGTGTC-----CTATCTGTCTTA 109
 QY 108 IleTyrSerAsp-ProGlyTyrHisSerGlnGluAspThrGluThr-----LysThr-S 125
 DB 110 TGTTCACAGAGTGCCTATTCACAGAGATCCAGATGACCAAAACCTCATCAAGACCAT 169
 QY 125 SerSerAsnGlyGluLysThrPheLeuTyrHisSerArgAlaIleThrGluValGluTyr 145
 DB 170 TGTTCACAGAGATCAATGACATTTTCACACAGCAGTGGTA-TCCGCCAAGCAGAGGCTCA 228
 QY 145 hrGlyLeuGlySerSerAlaGlyLeuValSerVal----- 156
 DB 229 CTGGCTTGACATTCATTCCTGGGCTTCACCCCATTCGATTTGTCACAGATGACCAAGA 288
 QY 157 -----ValAlaThrSerLeuLeuSerHis-----PheIleP 167
 DB 289 CTCTGGACGCTATCAACAGAGTCCACAGAGCTGCTCCCAAAATGCTGTCAGATGAG 348
 QY 167 roAsnValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysT 187
 DB 349 CCAATGACCTGGAGATCTCCAGAGCTCTCATCTGCTGCTTCTCCAAAGAGCTGT 408
 QY 187 yralGlnLysLysIleGly-----SerGlyPheAsp-----ValAlaThr 201
 DB 409 CCGTGGCTCAGACAGTGCCTGACAGAGCAGAGAGCCTGATGGCCGCTCGAAGCCT 468
 QY 201 lalleTyrGly-----LeuIleValTyrArgArgPheGlnProAlaLeuIleAsnAspV 219
 DB 469 CACTCACTCCACAGAGGTGTGTGCTTGTGACAGAGCTGACAGGCTCTCTG---CAGSACA 525
 QY 219 alPheGlnValLeuGlnSerAspProGlu-----L 229
 DB 526 TTTCTTCACAGTGTGATGTTAGCCCTGAATGCTGAAGTTTCAAGGCCACAGGCTCCCA 585
 QY 229 yAspProThrGluLeuLysLys-LeuIleGlnSerAsnTyrGluGluLysHisGluArg 248
 DB 586 AGAATCATGTAGAGGAGAAAGAAACCTTGCAGAGGCGCTTCAGAGAGAGAGAGACCATG 645
 QY 249 CysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlyLysSerGluThr 268
 DB 646 TCCACACATCCACATTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
 QY 269 ProLysLeuValSerArgValLeuGlnTyrLysLysGluLysProGluGlnSerSerVal 288
 DB 706 CCAATGCTTGAATGATTAATCAACACACTTATGACACAAAGAGAGGCGCAGC--- 762
 QY 289 ValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGluLeuArgGluMetArg 308
 DB 762 ----- 762
 QY 309 GluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHisSerValGluPro 328
 DB 763 -----CTCAGAGGCGGACTCTCACCT 783
 QY 329 LeuThrValAlaIleLysAsnIleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGlu 348
 DB 784 AGTTCTTACGAAAGTAGATTAAGAGCCATCCCATCCCTCATATGTCCCACTGCTCCGG 843

QY 349 ValProIleGluProAspValGlnThrGlnLeuLeuAspArgCysGlnGluIleProGly 368
 DB 844 GTACATGTCTCCCTGGGTACACGCTTGCCTGCGCCACGAGAGAGTGAAGTA----- 897
 QY 369 CysValGlyGlyValValProGlyAlaGlyTyrArgAlaIleAlaValLeuValLeu 388
 DB 898 -----GGATGGGTAGAGCCTTTGGGCTGTCTCAGAGCTTTGG 936
 QY 389 GluAsnGlnValGlyAsnPhe-----LysGlnLysThrLeuGlu 401
 DB 937 GAGCACCGTGAGAGCTGCATCCACACAGCTGGAATCTCCAAAGCAGACAGCATGGA 996
 QY 402 AsnProAspTyr---PheHisAsnValTyrTyrValAspLeuGluGlu 416
 DB 997 GCATTAATTAATTAATTCGATCTATTTGATGATGATGATGATGATGATGATGATGAT 1044

RESULT 26
 US-09-620-312D-338
 ; Sequence 338, Application US/09620312D
 ; Patent No. 6569662

GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom
 ; APPLICANT: Liu, Chenghua
 ; APPLICANT: Asundi, Vinod
 ; APPLICANT: Zhang, Jie
 ; APPLICANT: Ren, Feiyan
 ; APPLICANT: Chen, Rui-hong
 ; APPLICANT: Zhao, Qing A.
 ; APPLICANT: Weinman, Tom
 ; APPLICANT: Xue, Aidong J.
 ; APPLICANT: Yang, Yonghong
 ; APPLICANT: Wang, Jian-Rui
 ; APPLICANT: Zhou, Ping
 ; APPLICANT: Ma, Yunding
 ; APPLICANT: Wang, Duntui
 ; APPLICANT: Wang, Zhiwei
 ; APPLICANT: John Tillinghast
 ; APPLICANT: Drmanac, Radoje T.
 ; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
 ; FILE OF INVENTION: Polypeptides

FILE REFERENCE: 784CIP2B
 ; CURRENT APPLICATION NUMBER: US/09/620,312D
 ; CURRENT FILING DATE: 2000-07-19
 ; PRIOR APPLICATION NUMBER: 09/552,317
 ; PRIOR FILING DATE: 2000-04-25
 ; PRIOR APPLICATION NUMBER: 09/488,725
 ; PRIOR FILING DATE: 2000-01-21
 ; NUMBER OF SEQ ID NOS: 1105
 ; SOFTWARE: pc FL_genes Version 1.0
 ; SEQ ID NO 338
 ; LENGTH: 4295
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (153)..(3389)
 US-09-620-312D-338

Alignment Scores:
 Pred. No.: 0.926 Length: 4295
 Score: 96.50 Matches: 54
 Percent Similarity: 36.40% Conservative: 29
 Best Local Similarity: 23.68% Mismatches: 64
 Query Match: 4.33% Indels: 81
 DB: 4 Gaps: 13

US-10-069-062-7 (1-432) x US-09-620-312D-338 (1-4295)

QY 164 HisPheIleProAsnValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIle 183
 DB 1287 CATTTCATG-----TCCATCTGCGACCACTGCTCCAAATG 1322
 QY 184 AlaHisCysTyrAlaGlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyr 203

[illegible]


```

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 16995 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-82

Alignment Scores:
Pred. No.: 10.2 Length: 16995
Score: 96.00 Matches: 71
Percent Similarity: 38.01% Conservative: 70
Best Local Similarity: 19.14% Mismatches: 128
Query Match: 4.30% Indels: 102
Gaps: 17
DB: 4

US-10-069-062-7 (1-432) x US-08-961-527-82 (1-16995)
QY 50 ArgILeLysIleSerSerProGlnPheAlaAsnGlyGluTrpGluTyr---HisIleSer 68
DB 14109 CGTGTCCGCCCTCATGCGGCTGAGATTGTAATGACGCTTGATTCGGTCTATCGTGAC 14168
QY 69 SerAsnThrGluLysProArgGluValGlnSerArg-----IleAsnProPheLeuGlu 86
DB 14169 TTGATTCTGCGCCAGGTCGATTATTCAGAAAGACAGACAGCTGCTGCTCAGCTGTCA 14228
QY 87 AlaThrIlePheIleValLeuAlaTyrIle-----GlnPro 98
DB 14229 GAAATACAGACAGATTCACTGCTCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14288
QY 99 ThrGluAlaPhe-----AspLeuGluIleIleIleTyrSerAspProGlyTyrHis 115
DB 14289 AATGATGTCTATTGTTGGTCAAGATTAAACAGTATTGTCAATACAGGTCCC----- 14339
QY 116 SerGlnGluAspThrGluThrIleTyrSerSerAsnGlyGluLysThrPheLeuTyrHis 135
DB 14340 -----AATACAGGTGGGAAGACCATCATGCTCAAA 14369
QY 136 SerArgAlaIleThrGluVal---GluLysThrGlyLeu-----Gly 148
DB 14370 ACTCTGGGCTTGACACAGGTCATAGCGCCCACTGAGATTCGCCGATTTTAGACAGACGGA 14429
QY 149 SerSerAlaGlyLeuValSerVal----- 157
DB 14430 AGTCGTGGTGGTATTTTGAAGAAATCTTTGATGATTTGAGATGAGACGATCTATTGAG 14489
QY 158 ---AlaThrSerLeuLeuSerHisPheIleProAsnValIleSerThrAsnLysAspIle 176
DB 14490 CAGAGCTTGTCTACTTCTTACTGATGATGACCAATATCTGT-----GATATT 14537
QY 177 LeuHisAsnValAlaGlnIleAlaHisCysTyrAlaGlnLysValIleGlySerGlyPhe 196
DB 14538 CTTCGCAAGGTCAACCAACATCTCA---CTTTACTTTGGATGATGTTGGGGGCTGGTACT 14594
QY 197 Asp-----ValAlaThrAlaIle----- 202

```

```

DB 14595 GATCCCAAGAGGAGACGCCCTTCGATGCTATTTGAGACCTTCGCTGGCTCAA 14654
QY 203 -----TyrGlyLeu----- 205
DB 14655 ATCAAGACCATGGACGACCACTATTCAGAACTCAAGGCTTACGATTTGACAGACC 14714
QY 206 -----IleValTyrArgPheGlnProAlaLeuIleAsnAspValPheGlnVal 222
DB 14715 TTGTGCAAAATAGCCAGATGAGATTGATACGTGACACTTCGGCCGACCATGCTTT 14774
QY 223 LeuGluSerAspProGluLysPheProThrGluLeuLysLeuIleGluSerAsnTrp 242
DB 14775 ATGCAGGGGTGTTCTCGGCCA-----AGTATGCC 14804
QY 243 GluGluLysHisGluAlaGlySerThrLeuProTyrGlyIleLysLeuLeuMetGlyAspVal 262
DB 14805 TTGAAATTTGCCAAAGCTTACGCTATGT-----GAAGTTATCTGAGAGATGCC 14855
QY 263 LysGlyGlySerGluThrProLysLeuValSerArgValLeuGlnTrpLysLysGlyLys 282
DB 14856 AGTCAGACATGATCAGACAGCAATGACGTCATCATTCATTCATTCATTCATTCATTCAT 14915
QY 283 ProGluGluSerSerValValTyrAsp-----GlnLeuAsnSerAlaAsnLeuGln 299
DB 14916 ACGCTGAAAGCGCGAAAGCTTTGACATATCCGTGAGGTGAGACCAAGAAATCTCAAG 14975
QY 300 PheMetLysGluLeuArgGluLeuLysTyrAspSerAspProGluThrTyrIle 319
DB 14976 ATGAACCGGCGCTTAAATACTTACACAGAGCTTAACTGTAAGAAAGAACCGAGCTT 15035
QY 320 ---LysGluLeuAspHisSerValGluProLeuThrValAlaIleLysAsnIleArgLys 338
DB 15036 AACAGGCGCGCGAAGAGGCTGCTGAGATTGTGATATGAGGCTTAAAGTAAAGTACGAG 15095
QY 339 GlyLeuGlnAlaLeuThrGlnLysSerGluVal 349
DB 15096 ATTCTCAAAATATCTCACAGTAATCCCAATCC 15128

RESULT 30
US-09-107-532A-2010
Sequence 2010, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
CORRESPONDENCE ADDRESS:
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSER: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 893-5007

```



```

Db      9132 -GACGAAACAAAGATTTATTAGTAATATCAAGATTT----- 9170
Qy      314 pProgluThrTyrIleLeuSerValGluProLeuThrValAlaIleu 334
Db      9171 -----CTTTTAACCATCATCTTAAATATGATGACGAGATTTGCCA 9211
Qy      334 sAsnIleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAs 354
Db      9212 CGGAATAACAGCTTCTACCAACCACTGCTCAATCAAAACAAATTAATCTGATGAAAA 9271
Qy      354 pValGlnThrGlnLeuLeuAspArgCysGlnGlu-----11 366
Db      9272 AATCCAA---CAGGTAAATTTCACATGCTCTGGAACGATGATCACGGTTTACTGATGTT 9328
Qy      366 eProGlyCysValGly-----GlyValValProGlyValaGlyG1 379
Db      9329 TCAGAGAGATAGGTGTAAGACAGATTTGTTGACTTCGGAAAT-----GGAGCATGGG 9382
Qy      379 YTYrAspAlaIleAlaValLeuValLeuGlnValGlyAsnPhelysGlnLysTh 399
Db      9383 GTATATTGGTGTCTATTAAATATAATATCCATTCGATGTCAGACATAGGAGAAAG 9442
Qy      399 rIleuGluAsnProAspTyrPheHisAsnValTyrTrpValaAsp 413
Db      9443 TATGAAA-----TTATTTTGACAAAC 9464

RESULT 33
US-08-969-046-3
Sequence 3, Application US/08969046B
Patent No. 645762
GENERAL INFORMATION:
APPLICANT: Chiang, Vincent Lee C.
APPLICANT: Teal, Chung-Jui
APPLICANT: Hu, Wen-Jing
TITLE OF INVENTION: Genetic engineering of trees through
FILE REFERENCE: 881.003US1
CURRENT APPLICATION NUMBER: US/08/969,046B
CURRENT FILING DATE: 1997-11-12
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 1710
TYPE: DNA
ORGANISM: Populus tremuloides Michx.
NAME/KEY: CDS
LOCATION: (1)... (1710)
US-08-969-046-3

Alignment Scores:
Pred. No.: 0.354 Length: 1710
Score: 94.50 Matches: 81
Percent Similarity: 37.87% Conservative: 72
Best Local Similarity: 20.05% Mismatches: 150
Query Match: 4.24% Indels: 101
DB: 4 Gaps: 19

US-10-069-062-7 (1-432) x US-08-969-046-3 (1-1710)
Qy      32 SerSerArgMetHisAlaValIleThrPro----- 41
Db      349 GCTTCATGATGGTGCAGTATCATCAGTGAACCTTTTCACATCAAGTGAATAA 408
Qy      42 ---LysGlyThrSerLeuLysSerArgGlyLeuLysSerSerProGlnPheAlaAsn 60
Db      409 TTCAGCAATTCCTGCTTCTCGTGAACATGATATCACCCAGTCTCAATATGTGAAC 468
Qy      61 -----GlyLysProGluThrHisIleSerSerAsnThrGlnLysProArgGluValGln 78
Db      469 AAGCTTAGAGATAGTATGATCCAT-----GAAAGAACCAAAACCGGGGGA----- 516
Qy      79 SerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnPro 98

```

```

Db      517 -----GATTTCATAGTATATCACCATGATGACCG 546
Qy      99 ThrGlu---AlaPheAspLeuGluIleIleIleTyrSerAspProGlyTyrHisSerGln 117
Db      547 CCAGAACTGCTCTACATTTCAATGCTGTGCGAGCT-----ACCGAG 591
Qy      118 GluAspThrGluThrLysThrSerSerAsnGlyGluLysThrPheLeuTyrHisSerArg 137
Db      592 AGTGAATGCCAAGCTTTCATCTCCGAGATGATCCGTGGCATTTACATCTCTTCA 651
Qy      138 AlaIleThrGlnValGlnLysThrGlyLeuGlySerSerAlaGlyLeuValSerVal 157
Db      652 GCGACAAACAGGCTCCCAAAAGAGATGATGACCAACAGAGCTTGATTAACAAGTGG 711
Qy      158 AlaThrSerLeuLeuSerHisPheIleProAsnValIleSerThrAsnLysAspIleLeu 177
Db      712 GCTCAACAAGTGTATGAGAG--ATCCAAATTTATCTTGAACCAAGATGACGTTGTT 768
Qy      178 HisAsnValAlaGlnIleAlaHisCysTyrAlaGlnLysLysIle----- 192
Db      769 TTATGCTTTTACCTTTGTTTACATCTTTTATGAACAGCGTGTGTTATGCTCGTTG 828
Qy      193 -----GlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyrArgArg 210
Db      829 AGAGCCGGTCT-----GCTGTTCTTTATGCAAAAG 861
Qy      211 PheGlnProAlaLeuIleAsnAspValPheGln-----ValLeuGluSer 225
Db      862 TTTGAGATAGATCACTGCTAGCTCATTTCAAGAACACATGTTTCGGTTCGGCTGAG 921
Qy      226 AspProGluLysPheProThrGluLeuLysLysLeuIleGluSerAsnTrpGluGluLys 245
Db      922 GTGCCACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 981
Qy      246 HisGluArgCysThrLeu-----ProTyrGlyIleLysLeu----- 257
Db      982 TCGATCAGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1041
Qy      258 -----LeuMetGlyAspValLysGlyLysGlyLysSerGluThrProLys 270
Db      1042 AGAGCAGGGTTCCTCAAGCCATCTCGGACAGGGTTATGGATGACAGAGCCGGGCT 1101
Qy      271 LeuValSerArgValLeuGlnTrpLysLysGluLys-----ProGluGluSerSerValVal 289
Db      1102 GTGCTATCATGTGCTTACCTTTCAAGCACTTTCCACCAAGTCTGCGTCTGT 1161
Qy      290 TyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGluLeuArgGluMetArgGlu 309
Db      1162 GGAACGGTGTAGAAACGACAGCTCAAGTCATT----- 1197
Qy      310 LysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHisSerValGluProLeu 329
Db      1198 -----GACCTCGAGACCGCT--CGCTCTCTTGTTTACAAAC--CAACTGCTGT 1239
Qy      330 ThrValAlaIleLys-----AsnIleArgLysGly-----LeuGlnAlaLeu 343
Db      1240 GAATCTGCATCCGGATTCACCAATCATGAAAGATATTGAATGACGCGGAACGCCACG 1299
Qy      344 ThrGlnLysSerGluValPro-----IleGluPro 353
Db      1300 GCAAAACACATAGAGCTTGAAGGTTGGCTCCACATGAGATATAGTTATGTGACAGAC 1359
Qy      354 AspValGlnThrGlnLeuLeuAspArgCysGlnGlnIleProGlyCysValGlyGlyVal 373
Db      1360 GACGACGAATTTTCTATGTTGATGATGAGAAATCATAAATTCAAAGCTTCCAG 1419
Qy      374 ValProGlyAla 377
Db      1420 GTGCCCGACGCG 1431

RESULT 34
PCT-US93-06251-64

```

Sequence 64, Application PC/TUS9306251
GENERAL INFORMATION:
APPLICANT: Wickstrom, Eric and Rife, Jason P
TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
NUMBER OF INVENTIONS: Stereospecific Alkylphosphonates and Arylphosphonates
CORRESPONDENCE ADDRESSES:
ADDRESSER: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06251
FILING DATE: 19930630
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Digilio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8586
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 2840 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-64

Alignment Scores:
Pred. No.: 0.814 Length: 2840
Score: 94.50 Matches: 54
Percent Similarity: 39.69% Conservative: 48
Best Local Similarity: 21.01% Mismatches: 94
Query Match: 4.24% Indels: 61
DB: 5 Gaps: 10

US-10-069-062-7 (1-432) x PCT-US93-06251-64 (1-2840)
QY 116 SerGlnGluAspThrGluThrLysSerSerAsnGlyGluLysThrPheLeuTyrHis 135
DB 2012 TCAGACAGTCAGAGAGGCTCACAGTAGTGTAACCTTCAAAAACAGGCTCTATCCA 2071
QY 136 SerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAlaGlyLeuValSer 155
DB 2072 GATGCTCACTTAGAGAACAGAGAAATGATTTAAACAGTAGAGATTAATGATGAC 2131
QY 156 ValValAlaIleThrSerLeuSerHisPheIleProAsnValIleSerThrAsnLysAsp 175
DB 2132 CGTTTGATGATCATCATCTCAAAATAT-TCCTACAGTAA- 2170
QY 176 IleLeuHisAsnValAlaGlnIleAlaHisCysTyrAlaGlnLys-----LysIle 192
DB 2171 -----AGGAAATCTGAGTCTGCCACTTGCAACTTAGTCAGAGACATAAACAGTG 2221
QY 193 GlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyrArgArgPheGln 212
DB 2222 GGAATGGCTTGTGCTGCGCGCTCA-----TCT 2251
QY 213 ProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGluLysPheProThr 232
DB 2252 CCGCTTCTGTGTGAAGATGTC-----ATTGTGAGATGATTAAGGAAATCATGAGA 2305

QY 233 GluLeuLysLeuLeuIleGluSerAsnTrpGluGluLysGluValArgCysThrLeuPro 252
DB 2306 GAAGTAATAGAG-----ACTTATTTAAACACAGAGAAAACCTAACTTG----- 2350
QY 253 TyrGlyIleLysLeuLeuMetGlyAspValLysGlyGlySerGluThrProLysLeuVal 272
DB 2350 ----- 2350
QY 273 SerArgValLeuGlnTrpLysGlyGluLysProGluLysSerValValTyrAspGln 292
DB 2351 -----ATTTGCAAAAGAGACAACTTCAGATGAAATGTAATAATGTTG----- 2395
QY 293 LeuAsnSerAlaAsnLeuGlnPheMetLysGluLeuArgGluMetArgGluLysTyrAsp 312
DB 2396 -----AGTACTCAAAATCTATGAGAGAACTACTGGAACACAGACAGAAATTACAG 2446
QY 313 SerAspProGluThrTyrIleLysGluLeuAspHisSerValGluProLeuThrValAla 332
DB 2447 AAAGACTTGAAATCTTGCAAGATGAAACATGCTCAAAAGATGAAATTTATGTTGAA 2506
QY 333 IleLysAsnIleArgGlyGlyLeu---GlnAlaLeuThrGlnLys-----SerGluVal 349
DB 2507 CAGAAAGACTTAGAGAAAAAATTTGACAGATTAATGAAACAAAATGTACTGTGACTCA 2566
QY 350 ProIleGluProAspValGlnThrGlnLeuLeuAspArgCysGlnGluIle 366
DB 2567 AATTGAAAAAGACAAAGAGGCTGAAATGACAGACAGACTTGACAGACTG 2617

RESULT 35
US-09-107-532A-2059
Sequence 2059, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSER: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 2059:
SEQUENCE CHARACTERISTICS:
LENGTH: 1941 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO

ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...1941
SEQUENCE DESCRIPTION: SEQ ID NO: 2059:
US-09-107-532A-2059

Alignment Scores:
Pred. No.: 0.5
Score: 94.00
Percent Similarity: 33.88%
Best Local Similarity: 20.25%
Query Match: 4.22%
Length: 1941
Matches: 49
Conservative: 33
Mismatch: 82
Indels: 78
Gaps: 10

US-10-069-062-7 (1-432) x US-09-107-532A-2059 (1-1941)
21 GluProIleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThr 40
1168 GAACAGTATTTTACATTCTATACACTTCTTGAAGATCAGTTGAATAGACGATCCACA 1227
41 ProlyGlyThrSerLeuGlySerArgGlyLeuIleSerProGlnPheAlaAsn 60
1228 AAGACACACGCTGGCTCACTTACTCAGAAACAGACATACACGCAAGAGCTCACT 1287
61 GlyIuTrpGluTyrHisIleSerSerAsnThrGluLysProArgGluValGlnSerArg 80
1288 -----TGGGAAGTCATGCTGCTCGCAATACCA----- 1317
81 IleAsnProPheLeuGluAlaThrIlePheIleValIleuAlaTyrIleGln-----Pro 98
1318 -----GCTTACTTAAGTTGTTCCCA 1338
99 ThrGluAlaPheAspLeuGlu-----IleIleIle----- 108
1339 ACCAATTTTGGCGAATTAGAAAGTTCTAGTCGACGACATCACTGCTAATGTCATCGCA 1398
109 -----TyrSerAspProGlyTyrHisSerGlnIleuAsp 119
1399 AAAACGCAATCAACATTTCTGTCATATCTATGATCTTGGCTATTAATCTCTCAAGACAC 1458
120 ThrGluThrLysThrSerSer-----AsnGlyGluLysThrPheLeuTyr----- 134
1459 AAGAGTACATTAATCTCAAGTTTATGAGACGAAAGAAATAGTTTCAATGAGCTTAA 1518
135 -----HisSerArgAlaIleThrGluValGluLysThrGly 146
1519 GTCGTTGCTGTAGATGTTGTAGCTTATCAAGCCGCAATGATCATCAAGAAAGAGGT 1578
147 LeuGlySerSerAla-----GlyLeuValSerValAlaIleThrSer 160
1579 GTCGAATATGAAAGCCGCTGCGACAAAAGCTTCTGGAGACTGTAAATVCCGCTGAAGATTCC 1638
161 LeuLeuSerHisPheIleProAsn----- 168
1639 ATGTTGTGACACATATCTTATGACAAAGATGGAAGCTAAATGATGGGAACCA 1698
169 ----ValIleSerThrLysAsp-----IleLeuHisAsnValAlaGlnIleAlaHis 185
1699 GTCMAAGTACGAAATTTCAAGACGATCTTGATATATAAAGTTTCCACAGGTAAACAT 1758
186 CysTyrAlaGluLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeu 205
1759 ACTGTACTTTCAGTTACTTGTGCTGAGAGATTTACATAGTGCAATCTTTTGTGCTTA 1818
206 IleVal 207
1819 TGTATT 1824

RESULT 36
US-09-308-022-4
Sequence 4, Application US/09308022

Patent No. 6291654
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.
TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS
NUMBER OF INVENTIONS: PNEUMONIAE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSER: MEETING, RAASCH & GEBHARDT, P.A.
STREET: 119 No. 6291654th Fourth Street, Suite 203
CITY: Minneapolis
STATE: Minnesota
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/308,022
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/20586
FILING DATE: 12-NOV-1997
APPLICATION NUMBER: 60/029,444
FILING DATE: 12-NOV-1996
APPLICATION NUMBER: 60/038,086
FILING DATE: 18-FEB-1997
APPLICATION NUMBER: 60/059,368
FILING DATE: 19-SEP-1997
APPLICATION NUMBER: 60/062,473
FILING DATE: 16-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: MOETING, Ann M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110,00430101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3023 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-308-022-4

Alignment Scores:
Pred. No.: 1.04
Score: 94.00
Percent Similarity: 35.43%
Best Local Similarity: 19.35%
Query Match: 4.22%
Length: 3023
Matches: 77
Conservative: 64
Mismatch: 135
Indels: 122
Gaps: 17

US-10-069-062-7 (1-432) x US-09-308-022-4 (1-3023)
9 GlyLysAlaPheLeuAlaGlyGlyTyrLeuValLeuGluProIleTyrAspAlaTyrVal 28
164 GCGTGGACCTGTTGTAGCATGCTTATGATGACCTTTGTTTATGTTTCCATTTATGCA 223
29 ThrAlaLeuSerSerArg-----MetHisAlaValIleThrPro 41
224 GTTATTTTGTATTCAGATAGCTGAGAGGAAAGCTAATATACATGATTAATATCCCA 283
42 LysGlyThrSerLeuLysGluSerArgIleLysIleSerSerProGlnPheAlaAsnGly 61
284 AATGCA---AGCATTAAG----- 298
62 GluTrpGluTyrHisIleSerSerAsnThrGluLysProArgGluValGlnSerArgIle 81

Db	299	-----AGATAAATAACAAATTCGATT	319
Qy	82	AsnProPheLeuGlnAlaThrIlePheIleValLeuAlaTyrlleGlnProThrGluAla	101
Db	320	-----TATATACAG-----TTCAATTTGAAGTAAATAAGTAAAGTAAAGAAAA	364
Qy	102	PheAsnValGlnIleIleIleTySerSerProGlyTyrHisSerGlnIleAspThrGln	121
Db	365	ATATGAAAGGAAATTAACATGTTT	388
Qy	122	ThrIlyrThrSerSerAsnGlyGluIlyrThrPheLeuTyrHisSerGlnAlaIleThrGln	141
Db	389	-----GCATCAAAAAGCGAAGAAAGAAAGTACATTATTCATATTCGTAAATTT	433
Qy	142	ValGluIlyrThrGlyLeuGlySerSerAlaGlyLeuValSerValAlaIleThrSerLeu	161
Db	434	-----AGTATTTGAGTGTACGTAGTACCTGTTCGACGCTTT	469
Qy	152	LeuSer-----HisPheIleProAsnValIleSerThrAsnIlyAspIleLeu	177
Db	470	CTTATGCGAAGCTGTGTTTCATCGACAGAGAAACGAGGAAAGTACCCAGACGCACTTCT	529
Qy	178	HisAsnValAlaGlnIleAlaHisCysTyrAlaGlnIlyrLeuIleGlySerGlyPheAsp	197
Db	530	TCTATATATGCGCAAGACACACATGAGAAAGCTGCTAAACAAAGTCTGCATGATATAT	586
Qy	198	ValAlaIleThrAlaIleTyGlyLeuIleValTyrArgArg	210
Db	587	ATAGAAAAAATGTTGAGGAGATTCACATACAGATAGAGAAAAACATCCCAAAATGCGCC	646
Qy	211	-----PheGlnProAlaLeuIleAsnAspValPheGlnValLeu	223
Db	647	TTAAACATTAAGCTGAGCGCAATTTAAACGAAGTATTTGCGTAATTAATGTTTAAAGAA	706
Qy	224	GluSerAspProGluIlyrPheProThrGluLeuIlyrLeuIleGluSerAsnTPGlu	243
Db	707	GAGAACTCGAAGATCAGTTCGCCGTCAAGAAATTAAGAACAAAGTTAAGACCGACGTTT	763
Qy	244	GluIlyrHisGluArgCysThrLeuProTyrGlyIleLeuLeuLeuMetGlyAspValIlyr	263
Db	764	GAGAAAGTTTAAAAAAGATACATTGAAACACAGAGAAAAAGTA-----	805
Qy	264	GlyGlySerGlnThrProIlyrLeuValIserArgValLeuGlnThrIlyrIlyrGluIlyrPro	283
Db	806	-----GCGAAGCTTAAGAGAGAGGTTGAAGAGCTTAAGAAAAAAGCCGAGATCAAAAA	859
Qy	284	GluGluSerSerValValTyrAsp-----	291
Db	860	GAAAGAAAGATCGTCGTACCTACCCACCAACCAATACTTACAAAACGTTGAACCTTGAATTCCT	919
Qy	292	-----GlnLeuAsnSerAlaAsnLeuGlnPheMetIlyrGluLeuArgIleuMet	307
Db	920	GAGTTGCAATGTGAAGTTAAAGAGCGGAGCTTGAACCTAGTAAAAACAG-----	967
Qy	308	ArgGluIlyrTyrAspSerAspProGluIlyrTyrIleIlyrGluLeuAspHisSerValIleGlu	327
Db	968	---GAAGCTTAAGAAATCTCGAAACCGGCGCACATTTAACAGCAAAAGAGAAAGTTGAG	1022
Qy	328	ProLeu-----ThrValAlaIleIlyrAsnIle-----ArgIlyrGlyLeu	340
Db	1025	AGTAAAAAAGCTGAGGCTTACAAAGTTTGAAGAAACATTAAGACACAGATGTGTAAGAAAAAGCAGAA	108
Qy	341	GlnAlaLeuThrGlnIlyrSerGluValProIle---GluProAspValGlnThr	357
Db	1085	GAAAGAAAGCTAAACGAAAGAACACAGATGCTAGTTGAAGGAAGCTAAATGTACGACAT	1138

RESULT 37

US-08-416-603-1

; Sequence 1, Application US/0841603

; Patent No. 586780

GENERAL INFORMATION:

APPLICANT: Law, Marcus

APPLICANT: Hebard, Ledare

```

? APPLICANT: Reddick, Bradford B.
? TITLE OF INVENTION: Maize Chlorotic Dwarf Virus Genome and
? TITLE OF INVENTION: Uses Therefor
? NUMBER OF SEQUENCES: 11
? CORRESPONDENCE ADDRESS:
? ADDRESS: Saliwanchik & Saliwanchik
? STREET: 2421 N.W. 41st Street, Suite A-1
? CITY: Gainesville
? STATE: FL
? COUNTRY: USA
? ZIP: 32606
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30B
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/416,603
? FILING DATE:
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Lloyd, Jeffrey
? REGISTRATION NUMBER: 35,589
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 904-375-8100
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS: 1:
? LENGTH: 11832 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? HYPOTHEICAL: NO
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 435..10763
? OTHER INFORMATION: /note= "cDNA of MCDV-Tn genome"
US-08-416-603-1

Alignment Scores:
Pred. No.: 9.76 Length: 11832
Score: 94.00 Matches: 66
Percent Similarity: 37.09% Conservative: 59
Best Local Similarity: 19.58% Mismatches: 122
Query Match: 4.22% Indels: 90
DB: 2 Gaps: 16

US-10-069-062-7 (1-432) x US-08-416-603-1 (1-11832)
QY 133 LeuTyrHisSerArgAlaIleThrGluValGluYerThrGlyLeu-----Gly 148
DB 1587 CTTTATCATGCGCAATTGTATTAATCTCGACAAATCACTTGGAATACGTCACCTTTGGA 1646
QY 149 SerSerAlaGlyLeuValSerValAlaAlaThrSerLeuLeuSerHisPheIleProAsn 168
DB 1647 GGAATGATAGGG---GTGAACATTGCATGCATAGGACCATTTATGGAATTCACAAACGT 1703
QY 169 ValIleSerThrAsnLysAspIle-----LeuHisAsnValAlaGlnIleAlaHisCys 186
DB 1704 TTCTATATCGCGGCACTTCAGAAATAGTCCATGAAATATGTATTGGAGAGAGACAGATGC 1763
QY 187 TyrAlaGlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIle 206
DB 1764 CAACCTGAG---TCGACTTCAT----- 1784
QY 207 ValTyrArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAsp 226
DB 1785 -----GATGAAGAAATTCACAAAGCATGAGCA-- 1811
QY 227 ProGluLysPhePheProThrGluLeuLysLysLeuIleGluSerLeuTrp----- 242
DB 1812 ---GAAAGAGCGCATGCGGAAATTCAAAGCGTC---TCAAATTTGGGTTAGTGAATAT 1862

```

```

QY 243 -----GluGluLeuHisGluArgCysThrLeuPro 252
DB 1863 CTTGATAGTAAGAAGCTCATTCATAGTCAGTAAGGCTGAAGAAACTCGAGGA 1922
QY 253 TyrGlyIleLeuLeuMetGlyAsp---ValIleGlyGlySerGluThrProIleu 271
DB 1923 TTAGGTTTAAATCAGGTTTGGAGAGATTCCTCAAGGTCGTCGATTCGTCGAGCAGC 1982
QY 272 ValSerArgValLeuGlnIleThrPheGlyGlu----- 281
DB 1983 TTRACACAGGTTTTCATCTGGCCCATTCATCTTCGATTCATTCGCGCAAAAGCCACAGCT 2042
QY 282 -----LysProGluGluSerSerValValTyrAspGlnLeuAsnSerIleAsnLeu 298
DB 2043 GATTGCTGCTGAAGTAATAGTCATCATCAGTACGACAGAAATCTGTCTGTGATCCCT 2102
QY 299 GlnPheMetLeuGluLeuArgIleMetArgGluIleTyrAspSerAspProGluThrTyr 318
DB 2103 GAATTCACAGAAATATGACAGATTCACG----- 2132
QY 319 IleLeuGluLeuAspHisSerValGluProLeuThrValAlaIleLysAsnIleArgLys 338
DB 2133 ---AAAGAAACCAAGATGGGAATAGATCTTACAGATTCGATTCAGAAATATATCAGAA 2189
QY 339 GlyLeuGlnAlaLeuThrGlnIleLysSerGluValProIleGluProAspValGlnThrGln 358
DB 2190 GGGATTGACAGATTCACGCAAGATTCAGAAACAAATTTGAA-----CCA 2234
QY 359 LeuLeuAspArgCysGlnGluIleProGlyCys-----ValGlyGlyValValPro 375
DB 2235 ATGTGTGATCGAATTAAGACCCATCGAGACCAAACTCAAAAGCTTAAAGATTTAAACCT 2294
QY 376 GlyAlaGlyGly---TyrAspAlaIleAlaValLeuValLeuGlnAsnGlnValGlyAsn 394
DB 2295 GATGCTGGGGGATCAAAAGATAGTGAAGCTATGCGCCAAATTCAGCGCCATCAAGAT 2354
QY 395 -----PheLeuGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyrTrp 411
DB 2355 ATTAAGCTATTAAACAGCAGATGATGAATTAATGATGAATTA----- 2402
QY 412 ValAspLeuGluGlnIleThrGluGlyValLeuGluGluLysProGluAsp 428
DB 2403 ---GACCTGAGGATTCAGACGATCATCAAGAAATTCAGGACGATGAT 2450

RESULT 38
US-08-545-528D-1/c
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; Patent No. 6537773
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB193PI
; CURRENT APPLICATION NUMBER: US/08/545,528D
; PRIOR FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Alignment Scores:
Pred. No.: 5.87e+03 Length: 580073
Score: 94.00 Matches: 113
Percent Similarity: 34.04 Conservative: 63
Best Local Similarity: 21.86 Mismatches: 161

```

```

Query Match: 4.22% Indels: 180
DB: 4 Gaps: 29
US-10-069-062-7 (1-432) x US-08-545-528D-1 (1-580073)
QY 48 GluSerArgIleLysIleSerSerProGlnPheAlaAsnGluTrp-----GluTyr 65
DB 487740 GAACAGAAAGTTAGGTTCAAGT---GAACCTCCCAAAAGTAAGTAAGTATGATGAATA 487684
QY 66 HisIleSerSerAsnThrGluLysProArgGluValGlnSerArgIleAsnProPheLeu 85
DB 487683 ACTTTATTAAATGAATGATCCAAACCTCAAGAGATTCGAATATGAATTAATTCCTT 487624
QY 86 GluAlaThrIlePheIleValLeuAlaTyrIleGlnProThrGluAlaPheAsp----- 103
DB 487623 GAAGCT-----GAACCAAAATCATTTATTTATGAAAAA 487591
QY 104 LeuGluIleIleIleTyrSerAspProGlyTyr----- 114
DB 487590 ACAACATATGTTGTAAGAAAGTGAACACCTTCATTCACACTGACCTATCATCTGAATTA 487531
QY 115 -----HisSerGlnLysPheThrGluThrIleTyrThrSerSer-----Asn 127
DB 487530 GATTCTGTAATGATGTTGATTAATCATTAAGAACTAAACATACTGCTGTGAATTAAT 487471
QY 128 GlyIleLys-----ThrPheLeu-----TyrHisSerArgAlaIleThrGlu 141
DB 487470 CATGAAGAAATGTAATGATTAATTTATTAATTTAGATGTTTCGAAAAAGAACTCAAGAA 487411
QY 142 ValGluLysThrGlyLeuGlySerSerAlaGlyLeuVal-----SerValVal 157
DB 487410 CAACCAACACACACATGAAACCTGATAGTGTGATTCACCAATATCAATCAATAGTT 487351
QY 158 AlaThrSerLeuLeu---SerIlePheIleProAsnValIleSerThr---AsnLysAsp 175
DB 487350 GAAGACAGTTTCATGAAAGTCTGAACCTCAATGATTTTCATCTCAACAAAGAT 487291
QY 176 IleLeuHisAsnValAlaGlnIleAlaHisCysTyrAlaGlnLysLysIleGlySerGly 195
DB 487290 ACGCTTGATTCATTATGCTCA-----ACTCAAGAGGTTGAGACTTCTGAA 487246
QY 196 PheAspValAlaThrAlaIleTyrGlyLeuLeuValTyrArgArgPheGlnProAlaLeu 215
DB 487245 TCAAATGTATCCACAGCTGGAACAGAAACA-----AAATTAATTCACATCAAGAT 487195
QY 216 IleAsnAspValPheGlnValLeuGluSerAspProGluLysPheProThrGluLeuLys 235
DB 487194 GAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 487156
QY 236 LysLeuIleGluSerAsn-----TrpGluGluLysHisGluArgCysThrLeuProTyr 253
DB 487155 GAATATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 487120
QY 254 GlyIleLysLeuLeuMetGlyAspValLysGlyGlySerGluThr-----ProLys 270
DB 487119 -----CCAGACGAAAAAGAGATTCGATTCAGAACTCAACTACT 487078
QY 271 LeuValSerArgValLeuGlnIleThrLysGlyGlySerProGluGluSerSer----- 287
DB 487077 TTCAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 487018
QY 288 -----ValValTyrAspGlnLeuAsnSerIleAsnLeu 298
DB 487017 GCTACTTTGATTAATTCATTCACACAGACAGATTTTGAATGAATTAATTAATTAATTAATTA 486958
QY 299 Gln-----PheMetLys 302
DB 486957 CAGCTGAAGTACTACTGAAGTTGTTTGTATGATCATTTTCAACAGATTCACACT 486898
QY 303 GluLeuArgGluMetArgGlyLysTyrAspSerAspProGlu-----ThrTyrIleLys 320
DB 486897 GAACAAACTCTCAAGAAAGCTAAATTTGATTCACAGCTGCAAGATTTCCCAAGAAATCATCG 486838

```

```

QY      321  GluLeuAspHisSerValGluProLeuThrValAlaIleIysAsnIleArgLys----- 338
      486837  CAAGCTGAATTTCATGCTGAACAATCTCTGAGAAATTAAATTAGAGAAAAAACTGAA 486778
QY      339  -----GlyLeuGlnAlaLeuThrGlnLysSerGluValProIle----- 351
      486777  GCAGTTTTCATCATCAACAATTAGAAAAATCAGCTCGAAGAAACAGTTGTACACCTACT 486718
QY      352  -----GluProAsp-----ValGlnThrGlnLeu----- 359
      486717  GAAGTTACTGCTTTTAAACCAAGAACATTGAAACCCAAATTAGAACCTTACCTGAAAGAT 486658
QY      360  -----LeuAspArgCysGlnGlnIleProGlyCysValGlyGlyVal 373
      486657  CAACATCTGAACCTGATTCAGT-----CAAAATCATCCAGAAATTTGTTACTGCTGAA 486604
QY      374  ValProGlyValAlaGlyIleValAspAlaIleValValLeuValLeuGlnValGly 393
      486603  GTTGAACAATC-----TTTGATGCAACTAGTTAGAGATTAAACCTAGAGAAAGCT 486550
QY      394  AsnPro----- 395
      486549  AACTTGATTAATGTTGAAAAATTAATGAAGTTCAACTTAAGAAACAGAACTGAATTTACT 486490
QY      396  -----LysGlnLysThr----- 399
      486489  TTGTAGAAACTTAAGAACTTCAACAGAAACCTCCCTGACCATTTATGACAGAAAGA 486430
QY      400  LeuGlnAsnProAspTyrPheHisAsnValTyrTrpValAsp----- 413
      486429  TTTAAATCTGACAGCTACTTTTGATTAATGTATCTGAAGCTGAATCAGAGAGATTTTTGA 486370
QY      414  -----LeuGlnGlnGlnThrGlnGlyValLeuGlnGlnLysProGlu 427
      486369  AAACCTCAATTAGAAACACAGCTGAAAAAATATATAGAGAGAACTTAA 486319

RESULT 39
US-09-328-352-1259
; Sequence 1259, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328.352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1259
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1259

Alignment Scores:
Pred. No.: 0.146 Length: 843
Score: 93.50 Matches: 63
Percent Similarity: 38.64% Conservative: 39
Best Local Similarity: 23.86% Mismatches: 116
Query Match: 4.19% Indels: 46
DB: 4 Gaps: 11

US-10-069-062-7 (1-432) x US-09-328-352-1259 (1-843)
QY      109  TyrSerAspProGlyTyrHisSerGlnGluAspThrGluThrLysThrSerSerArgGly 128
      112  TACAGTCACAGATTAATTATTCATAGCATTAAGTAAATTAGAACATTTTCCTCAATTAATCCT 171
QY      129  GlnLysThrPheLeuTyrHisSerArgAlaIleThr-----GluValGlnLys 144
      172  GATAGTACTTCTTGCCATATGCAAAAGATTAATGATTCATTAAGCGTGAAGAGTGAATC 231
QY      145  ThrGlyLeuGlySerSerAlaGlyLeuValSerValAlaThrSerLeu----- 162

```

```

DB      232  AACGGTCTAAAAAATGGCATTTGTCACAAAGTTGATGATCATATGATGAATTGAA 291
QY      163  SerHisPheIleProAsnValIleSerThrAsnLysAspIleLeuHisAsnValAlaGln 182
DB      292  CTGCATTAACCGTGTCAAAAGTATTTACTTCAATTAAGATTAACCAAAATTTTGTGAC 351
QY      183  IleAlaHisCysTyrAlaGlnLysLysIleGlySerGlyPheAspValAlaThrAlaIle 202
DB      352  ACTGAGAGAGCATATCTT-----GTAGACCTTGCTTTATCATACATACGCTGCTCA 402
QY      203  TyrGlyLeuIleValTyrArgArgPheGlnProAlaLeuIleAsnAspValPheGlnVal 222
DB      403  TTAGT-----ACTCAGATCATGAGGTTCC 426
QY      223  LeuGlnSerAspProGlnLysPheProThrGlnLeuLysLeuIleGlnSerAsnTrp 242
DB      427  GCATCAAAAATTCACAGAACCAACCATTTGTTGTTAAACAGATTAATAATTA 486
QY      243  GluGlnLysHisGluArgCysThrLeuProTyrGlyIleLeuLeuLeuMetGlyAspVal 262
DB      487  TTACTGCTGCTGAGAAAGCTGATCTCAATT----- 519
QY      263  LysGlyGlySerGlnThrProLys-----LeuValSerArgValLeuGlnTrpLys 280
DB      520  GAAGATGCCCAAGATGACACAGCAAAAGGCTGTTTCTTAAAGTTTGATTTGATTGAAG 579
QY      281  GluLysProGlnLysSerSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPhe 300
DB      580  CCACCAACAAGAACAA-----CAAGAAATCAAAATTAAGAACCAATTC 621
QY      301  MetLysGlnLeuArgGlnMetArgGlu-----LysTyrAspSerAspProGlnTrp 317
DB      622  ---AAGAACTTGCACATCTGAGAGGATTCCTTAACATTTGCTGAATAATCAAC 678
QY      318  TyrIleLysGlnLeuAspHisSer-ValGlnProLeuThrValAlaIleLysAsnIleArg 337
DB      679  AAGTTGAAAAGGTCATACAGTATTCAGAACTGCAATCAACAGCATTTAAATTGAA 738
QY      337  GlySerGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGlnProAspValGlnTrp 357
DB      739  AAG-----AATTCATATACCTTAAAGCCAAAGCTTGAAGGTGAAG---AAATCCACAG 789
QY      357  rglnLeuLeu 360
DB      790  AACACCTTCTG 799

RESULT 40
US-09-120-663-1
; Sequence 1, Application US/09120663
; Patent No. 6228644
; GENERAL INFORMATION:
; APPLICANT: Bogdanove, Adam J.
; APPLICANT: Kim, Jihyun Francis
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,663

```


FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/055,105
 FILING DATE: 06-AUG-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Goldman, Michael L.
 REGISTRATION NUMBER: 30,727
 REFERENCE/DOCKET NUMBER: 19603/1661
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (716) 263-1304
 TELEFAX: (716) 263-1600
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5517 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-09-120-663-1

Alignment Scores:
 Pred. No.: 3 2
 Score: 93.50
 Percent Similarity: 37.84%
 Best Local Similarity: 20.55%
 Query Match: 4.19%
 DB: 3
 Gaps: 19

US-10-069-062-7 (1-432) x US-09-120-663-1 (1-5517)

QY 36 HisAlaValIleThrPro-----LysGlyThr-----SerLeuYsglu 48
 DB 4246 CATTGGACCTTGACCCCGCAGAACTGTTGCCAAAGGGATGCAACATGATGAGCG 4305
 QY 49 ---SerArgIleYsleSerSerProGlnPheAlaAsnGlyGluTrpGluTyrHisIle 67
 DB 4306 GCGAGCAACTGACGTTAGCTGATCCTCGCAATCTGATCTCGCGATCTCGCGGTATC 4365
 QY 68 SerSerAsnThrGlu-----LysProArgGluValGlnSerArgIleAsnProPheLeu 85
 DB 4366 AATCTGAACGAAAGCGGACGTAACCAATGCTGCTACTGCGGTGTTCTCCGCGCTA 4425
 QY 86 GluAlaThrIlePheIleValIleuAlaTyrIleGlnProThrGluAlaPheAspLeuGlu 105
 DB 4426 AGTGCATCGCAAACTGCGCCGCGCTGCGTGAACGACACACACC----- 4473
 QY 106 IleIleIleTyrSerAspProGlyTyrHisSerGlnGluAspThrGluThrHisIle 125
 DB 4474 -----TCGGCCAGTTTGGCAGCAGCACTTCGGCC 4503
 QY 126 SerAsnGlyGluThrPheLeuTyrHisSerArgAlaIleThrGluValGluYsThr 145
 DB 4504 AGCAATTAACCGCCCACTTCTC-----AAC 4530
 QY 146 GlyLeuGlySerSerAlaGlyLeuValSerValAlaIleThrSerLeuLeuSer----- 163
 DB 4531 GGGGTCGGCGCGGCTTAACTGACGCGCTCTTAGGGGTTGCCATTCACTACGCAAT 4590
 QY 164 -----HisPheIleProAsnValIleSerThrAsnYsAspIleLeuHis 178
 DB 4591 GAAGGAAACCGGTCGATCTTCCGCGCATTTACTGACCAATGTTTC----- 4641
 QY 179 AsnValAlaGlnIleAlaHisCysTyrAlaGlnYsIleGlySerGlyPheAspVal 198
 DB 4642 -----GCGCGCTGGCG-----CTGATTAAC 4662
 QY 199 AlaThrAlaIleTyrGlyLeuIleValTyrArgArgPheGlnProAlaLeuIleAsnAsp 218
 DB 4663 CGTACTCTCAGAGATATGAGCTGGAATGAAAGCGCGGACCGGTGACCAAGCAAGAT 4722
 QY 219 ValPheGlnValLeu-----GluSerAspProGluYsPhe 230

DB 4723 ATCAGCGATTGACCTCCAGCGTGGAAAAACACTTAAAGATAGGCCCAACAGAAATG 4782
 QY 231 ProThrGluLeuYsLeuIleGluSerAsnTrpGluYsHisGluArgCysThr 250
 DB 4783 CTGGCGCTCTCAAGAGATTAGATACCGCTAAGCCCGCTGAACCACTGATTTTACAG 4842
 QY 251 LeuProTyrGlyIleYsLeuLeuMetGlyAsp-----ValYsGlyYsSerGluThr 268
 DB 4843 CAGCATTTCAAGTCAAAAGATGTGTCGTGATGAACCTACAGAGCGGTGGCAACTG 4902
 QY 269 ProYsLeuValSerArgValLeuGlnIleTyrYsGluYsProGlu----- 284
 DB 4903 AAAAACTGGTGAACGTCACAGCGCTGCGACAGCCACAGCATGGAATTAGATCTGCC 4962
 QY 285 GluSerSerValIleTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetYsGluLeu 304
 DB 4963 AGTCACAGCAGACCTTACATTAATCTTGCAGAAATTAATATGACGGCATTTGCGAGCTG 5022
 QY 305 ArgGluMetArgGluYsTyrAspSerAsp---ProGluThrTyrIleYsGluLeuAsp 323
 DB 5023 -----CTACACAAACATTTGATGCGCATTTACAGCAGACAGATGCCAAACGTTTGGT 5076
 QY 324 HisSerVal-----GluProLeuThrValAlaIleYsAsnIleArgYsGlyLeuGln 341
 DB 5077 GAATGATGAATTAACGATCCG-----GCATGAAAGATATTATTAAAGCAGCTGCA 5127
 QY 342 AlaLeuThrGlnYsSerGluValProIleGluProAspValGlnThrGlnLeuLeuAsp 361
 DB 5128 AGT---ACGCGCTTACAGAGCGCCAGCGTGTGATGAGCTGAAAGATGCTGTGCTAG 5184
 QY 362 ArgCysGlnGluIleProGlyCysValGlyGlyValAlaProGlyAlaGlyYsTyrAsp 381
 DB 5185 CAGACGGAATA-----GCATTAATCGACGCGTAAGCTGCTGTGA 5226
 QY 382 AlaIleAlaValLeuValLeuGluAsnGlnValGlyAsnPheYsGlnYsThrLeu 400
 DB 5227 GAAGTGGAGTAACTTTTCAGAGATCGT-----AACAACTTGGCGGTAAATCGGTC 5277

Search completed: August 15, 2003, 15:56:36
 Job time : 986 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 15, 2003, 12:39:53 ; Search time 360 Seconds
(without alignments)
3239.326 Million cell updates/sec

Title: US-10-069-062-7
Perfect score: 2230
Sequence: 1 MSKAFSAPGKAFIAGYIVL.....DLEQTEGVLEKPEPDYICL 432

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/sgn2.1/USFPO/spool/US10069062/runat.07082003.161502.4026/app.query.fasta_1.583
-DB=N.Geneseq_19Jun03 -QFMT=fasta -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=pcio -NORF=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10069062.GCGN 1.1.312 @runat.07082003.161502.4026 -NCPU=6 -ICPU=3
-NO MMAP -LARGEBUFFER -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: N.Geneseq_19Jun03:*
2: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1980.DAT:*
3: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1981.DAT:*
4: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1982.DAT:*
5: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1983.DAT:*
6: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1984.DAT:*
7: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1985.DAT:*
8: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1986.DAT:*
9: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1987.DAT:*
10: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1988.DAT:*
11: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1989.DAT:*
12: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1990.DAT:*
13: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1991.DAT:*
14: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1992.DAT:*
15: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1993.DAT:*
16: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1994.DAT:*
17: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1995.DAT:*
18: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1996.DAT:*
19: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1997.DAT:*
20: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1998.DAT:*
21: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA1999.DAT:*
22: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA2000.DAT:*
23: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA2001A.DAT:*
24: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA2001B.DAT:*
25: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA2002.DAT:*
26: /SIDSI/gcgcdata/geneeq/geneeqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2222	99.6	1299	22	AAD02792	Candida albicans p
2	2222	99.6	1763	22	AAD02791	Candida albicans p
3	2208	99.0	1299	22	AAS23422	Candida albicans e
4	2208	99.0	1299	24	ABZ31736	Candida albicans e
5	877	39.3	577	22	AAD02788	S. cerevisiae ERG8
6	819	36.7	547	22	AAD02787	S. cerevisiae ERG8
7	809.5	36.3	1356	24	AAL40805	Nucleic acid relat
8	809.5	36.3	1356	24	ABK96801	S. cerevisiae pren
9	809.5	36.3	1356	24	AAD31012	Yeast phosphomeryl
10	809.5	36.3	7681	24	AAD31026	Operon D DNA encod
11	809.5	36.3	7693	24	AAD31023	Operon A DNA encod
12	809.5	36.3	7695	24	AAD31024	Operon B DNA encod
13	809.5	36.3	8224	24	AAD31027	Operon B DNA encod
14	809.5	36.3	8235	24	AAD31025	Operon C encoding
15	809.5	36.3	8400	24	AAD31029	Operon G DNA encod
16	809.5	36.3	13917	24	AAD31037	Plastid transforma
17	809.5	36.3	14623	24	AAD31039	Plastid transforma
18	809.5	36.3	14623	24	AAD31041	Plastid transforma
19	671	30.1	1356	25	ABT19236	Aspergillus fumiga
20	655.5	29.4	1455	25	ABT21056	Aspergillus fumiga
21	649.5	29.1	1509	25	ABT18642	Aspergillus fumiga
22	649.5	29.1	1509	25	ABT20458	Aspergillus fumiga
23	649.5	29.1	3508	25	ABT18048	Aspergillus fumiga
24	649.5	29.1	3509	25	ABT19862	Aspergillus fumiga
25	475	21.3	2396	24	ABL60244	Arabidopsis thalia
26	445.5	20.0	1509	25	ABL55372	Acetyl-CoA acetyl
27	388.5	17.4	869	25	ABZ51452	Aspergillus oryzae
28	271	12.2	586	21	AAF08722	Fusarium venenatum
29	231	10.4	728	24	ABL60246	Gossypium hirsutum
30	213	9.6	5158	25	ABT19456	Aspergillus fumiga
31	209.5	9.4	1107	22	AAF27596	Mevlonate pathway
32	209.5	9.4	1107	23	AAS52922	Enterococcus faeca
33	209.5	9.4	6835	20	AAK13141	Enterococcus faeca
34	209.5	9.4	6835	24	ABK98936	Enterococcus faeca
35	193	8.7	1077	22	ABF27619	Mevalonate pathway
36	181.5	8.1	757	24	ABO82712	Rice mevalonate ki
37	181.5	8.1	1077	22	AAH52762	S. epidermidis ope
38	181.5	8.1	1077	22	AAF27620	Mevalonate pathway
39	181.5	8.1	1089	24	ABN33055	Staphylococcus epi
40	181.5	8.1	4010	22	AAH54932	S. epidermidis gen
41	177.5	8.0	1086	22	AAF27618	Mevalonate pathway
42	173	7.8	1035	23	AAS52217	Staphylococcus aur
43	173	7.8	1077	22	AAF27601	Mevalonate pathway
44	173	7.8	1077	23	AAS55131	Staphylococcus aur
45	170.5	7.6	2944528	24	ABR03041	Listeria monocytos

ALIGNMENTS

RESULT 1	ALIGNMENTS
AAD02792	
ID AAD02792 standard; DNA; 1299 BP.	
XX	
AC AAD02792;	
XX	
DT 31-MAY-2001 (first entry)	
XX	
DE Candida albicans phosphomevalonate kinase (PMK) coding sequence.	
XX	
DE Phosphomevalonate kinase; PMK; ERG8; anti-fungal agent; diagnosis;	
KW infection; ds.	
XX	
OS Candida albicans.	
XX	
Key Location/Qualifiers	
FT CDS 1..1299	

XX XX Gene identification; essential gene; GRACE; pathogenic fungus;
 KW gene replacement and conditional expression; fungal infection; de.
 XX Candida albicans.
 XX WO200160975-A2.
 XX 23-AUG-2001.
 PD 20-FEB-2001; 2001WO-US05551.
 PF 18-FEB-2000; 2000US-0183534.
 PR (ELIT-) ELITRA PHARM INC.
 PA Roemer T, Jiang B, Boone C, Bussey H;
 PI WPI; 2001-489080/53.
 DR P-PSDB; AAU15094.
 DR
 XX
 PT Identifying genes essential to fungal metabolisms and identifying
 PT potential therapeutic agents that target these genes -
 PS Claim 22; Page 162; 324pp; English.
 CC The present invention relates to novel methods for constructing fungal
 CC strains useful for identification and validation of gene products as
 CC targets for therapeutic agents, for creating a collection of identified
 CC essential genes, and screening assays for the discovery of new drugs.
 CC The invention provides the GRACE (gene replacement and conditional
 CC expression) method for the construction of mutant organisms referred to
 CC as GRACE strains of the organism. The invention can be applied to any
 CC organism, particularly a pathogenic fungus e.g. *Candida albicans*,
 CC *Aspergillus fumigatus* and *Cryptococcus neoformans*. The methods are
 CC useful to identify agents that may be used in the treatment of fungal
 CC infections. AAS23381-AAS23442 represent *C. albicans* essential genes.
 CC
 SQ Sequence 1299 BP; 439 A; 221 C; 271 G; 368 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,126-218 Length: 1299
 Score: 2208.00 Matches: 428
 Percent Similarity: 99.54% Conservative: 2
 Best Local Similarity: 99.07% Mismatches: 2
 Query Match: 99.01% Indels: 0
 DB: 22 Gaps: 0
 US-10-069-062-7 (1-432) x AAS23422 (1-1299)
 QY 1 MetSerIyAlaPheSerAlaPProGIyLyVaAlaPheLeuValaGIyGIyTyrLeuValleu 20
 DB 1 ATGTCAAAAGCATTATGTGCACTGGAAGACATTTCTGCTGTGTGATATTTGGTCTT 60
 QY 21 GIUProIleTySerAaIaTyraIThrAlaLeuSerSerArgMetHisAlaValIleThr 40
 DB 61 GAGCCATTATTAATGCTTATGTGACAGCACTTCTCATCAAGATGATGAGATTATTAACA 120
 QY 41 ProLySGIyThSerLeuLySGIuSerArgIleLyIleSerSerProGInPheAlaen 60
 DB 121 CCAAAAAGGAACCGITTTGAAGAAATCTAGATCAAAATTTCTTCACCCCAATTTGCAAA 180
 QY 61 GIyGIuTProGIyThIleSerSerArgThGIuLyProArgGIuValGIuSerArg 80
 DB 181 GGAAGATGGAAATATCATATCATCAATATCAAGAAACCCAAAGAGTTCAGTACGC 240
 QY 81 ILeaSProPheLeuGIuAlaThrIlePheIleValIleuAlaTyriIleGIuProThrGIu 100
 DB 241 ATAAATCCATTTTATAGGCACTATATTCATGTTTATTTCAACCGACGAA 300
 QY 101 AlaPheAPLeuGIuIleIleIleIleTySerAPProGIyTyriHiserGIuAPThr 120
 DB 301 GCATTTGATCTTGAAATCATATTATTTACTGGAACCTGGATATCATCAAGAAAGATACT 360
 pb

QY 121 GIUThrIleTySerSerAsnGIyGIuLyThrPheLeuTyriHiserArgAlaIleThr 140
 DB 361 GAACCCAGACATCTCTCGAATGAGAAAATTTTCTTTACATCTCTGTCCTTAC 420
 QY 141 GIuValGIuLyThrGIyLeuGIySerSerAlaGIyLeuValSerValAlaThrSer 160
 DB 421 GAAGTGAAGAAACCGGATTAAGTTCATCGCAGAGATTAGTCTCAGTTGTTCCACAG 480
 QY 161 LeuLeuSerHisPheGIleProAsnValIleSerThrAsnIyAsnIleuHisAsnVal 180
 DB 481 TTATTAATCCATTTTATATCCCAATGTATCACTACGAATTAAGATATTTTGCAACAG 540
 QY 181 AlaGIuIleAlaHisCyTyraIaGIuLyLyIleGIySerGIyPheAspValAlaThr 200
 DB 541 GCACAGATTGCACATTTGTTATGCCAAAAGAAAGATGATCTGGGTTGATTTGCCACT 600
 QY 201 AlaIleTyrgIyLeuIleValIyraIrgArgPheGIuProAlaLeuIleAsnAspValPhe 220
 DB 601 GCAATTTATGCTGATGTATATAGAAATTTACAGCCAGCTTGTATTAATGACCGTGT 660
 QY 221 GIuValIleuGIuSerAPProGIuLyPheProThrGIuLeuLyIleGIuSer 240
 DB 661 CAGGTTCTAGAAAGATCTCGAAGAGTCCCAAGAGTTGAAAGAAATTTGATTCAGT 720
 QY 241 AenTPGIuGIuLyHisGIuArgCyThrIleuProTyGIyIleIySerLeuMetGIy 260
 DB 721 AACTGGGAATTCAAACATGAAAGATGTACTATTCACACAGGAATCAAGTTATTAAGGT 780
 QY 261 AspValIyGIyGIySerGIuThrProIySerValSerArgValIleuGIuThrIyIleTy 280
 DB 781 GACGTCAAGGGGTGCTCAGAAACACCAAAATGTGATACGAGTCTCAATGGAAGAA 840
 QY 281 GIUySProGIuLySerSerValIyTyraSPGIuLeuAsnSerAlaAsnIleuGIuPhe 300
 DB 841 GAAAGCCAGAAAGAGCTGTGTGTGTGATGACCACTTAATAGTCCATTTACAGTTT 900
 QY 301 MetIySGIuLeuArgIuMetArgIuTyraSPSerAPProGIuThrTyriIleTy 320
 DB 901 ATGAAGGAATTAAGGGAATGCGTGAATAATCGATCGATCCAGAGACTTATATTAA 960
 QY 321 GIULeuAspHisSerValGIuProLeuThrValAlaIleIyAsnIleArgIyGIyLeu 340
 DB 961 GAGTTAGATCATTTCTGTGAGCCCTTGTGACTGTGGATTAAGAAATCAGAAAGGGTTA 1020
 QY 341 GIuAlaLeuThrGIySerGIuValProIleGIuProAspValGIuThrGIuLeu 360
 DB 1021 CAAGCATTAACCAAAATCAGAGGTTCCAATTTGAACCTGATGTCCAAACCAAGTTGTG 1080
 QY 361 AspArgCySGInGIuIleProGIyCySerValGIyGIyValIaIleGIyGIyTyr 380
 DB 1081 GACCGTTGCAAGAAATCTCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
 QY 381 AspAlaIleAlaValIleuValIleuGIuAsnValIleGIyAsnPheIySGInLyThrLeu 400
 DB 1141 GATGCAATATGCTGTATGTGTGAAATATCAAGGGGAAATTTTAAGCAGAAATCTCTT 1200
 QY 401 GIuAsnProAspTyriPheHisAsnValTyriTPValaAPLeuGIuGIuThrGIuGIy 420
 DB 1201 GAAATATCCAGATTAATTTTCAATATGTATTAATGTGGTGAAGAGCAAGAGAGGT 1260
 QY 421 ValIleuGIuGIuLyPProGIuLyAPTyriIleGIyLeu 432
 DB 1261 GTACTTGAAGAAACCAAGAAACATATATAGTTTA 1296
 DB
 RESULT 4
 ID AB231736 standard; DNA; 1299 BP.
 AC AB231736;
 AC XX
 AC XX
 DT 30-JAN-2003 (first entry)
 XX

DE Candida albicans essential gene SEQ ID NO 6023.
 XX Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
 XX signal transduction; DNA replication; cell division; growth;
 XX proliferation; Candida albicans; fungicide; antifungal; gene; ss.
 OS Candida albicans.
 XX
 XX WO200253728-A2.
 XX
 XX 11-JUL-2002.
 XX
 XX 26-DEC-2001; 2001WO-US49486.
 XX
 XX 29-DEC-2000; 2000US-259128P.
 XX 20-FEB-2001; 2001US-079202A.
 XX 22-AUG-2001; 2001US-314050P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX
 XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL,
 XX
 XX MPI: 2002-566694/60.
 XX P-PSDB; ABE73186.
 XX
 XX Constructing strains for identifying gene products as effective targets
 XX for therapeutic intervention, by inactivating in the strain one allele
 XX of a gene and placing other allele of the gene under conditional
 XX expression -
 XX
 XX Claim 37; SEQ ID NO 6023; 167pp + Sequence listing; English.
 XX
 CC The invention relates to constructing (M1) a strain of diploid fungal
 CC cells in which both alleles of a gene are modified, comprising modifying
 CC one allele by insertion or replacement by a cassette having an
 CC expressible selectable marker and modifying other allele by
 CC recombination, of a promoter replacement fragment with a heterologous
 CC promoter, so that expression of the second allele is regulated by the
 CC promoter. (M1) is useful for constructing a strain of diploid fungal
 CC cells in which both alleles of a gene are modified. The diploid fungal
 CC cells having both alleles modified are useful for identifying a gene that
 CC is essential to the survival or growth of a fungus, a gene that
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene
 CC that contributes to the resistance of a diploid fungus to an antifungal
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus
 CC and for identifying a therapeutic agent for treatment of a mammalian
 CC disease. (M1) is useful for identifying a compound which modulates the
 CC activity of a gene product, preferably enzymatic activity, carbon
 CC compound catabolism, signal transduction, transporter, transcriptional,
 CC translational, signal transduction, DNA replication and cell division
 CC activity. The method is useful for identifying a compound having the
 CC ability to inhibit growth or proliferation of C. albicans cells and for
 CC treating infection by C. albicans. The present sequence is that of an
 CC essential Candida albicans gene used in the method of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office.
 XX
 XX SQ Sequence 1299 BP; 439 A; 221 C; 271 G; 368 T; 0 other;
 XX
 XX Alignment Scores:
 XX Pred. No.: 2,12e-218 Length: 1299
 XX Score: 2208.00 Matches: 428
 XX Percent Similarity: 99.54% Conservative: 2
 XX Best Local Similarity: 99.07% Mismatches: 2
 XX Query Match: 99.01% Indels: 0
 XX DB: 24 Gaps: 0
 XX
 XX US-10-069-062-7 (1-432) x AB231736 (1-1299)
 QY 1 MetSerLysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyGlyTyrLeuValIleu 20
 Db 1 ATGTCAAAAGCAATTAGTGCACCTGGAAAAAGCAATTTCTTGCTGTGGATATTGTTCTT 60

QY 21 GIuProIleTyAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThr 40
 Db 61 GAGCCAAATTTAGATGCTTATGACAGATTTGTATACAGAAATGATCAGATTATTAACA 120
 QY 41 ProLysGlyThrSerLeuLysGluSerArgIleLysIleSerSerProGlnPheAlaAsn 60
 Db 121 CCAAAAGCAACAGTTTGAAGAAATCTAGAAATCAAAATTTCTTCAACCCCAATTTGCAAAAC 180
 QY 61 GlyIuTPGluTyrHisIleSerSerAsnThrGluLysProArgGluValGlnSerArg 80
 Db 181 GGAGAAATGGAAATATCAATATCATCAATACAGAAAAACCAAGAAAGTTCAATGACAGCC 240
 QY 81 IleAsnProPheLeuGluAlaThrIlePheIleValIleuAlaTyrIleGlnProThrGlu 100
 Db 241 ATAAATCAATTTTGAAGCAACTATATCATCTGTTTAGCTTATATTCACCAAGCCGAA 300
 QY 101 AlaPheAspLeuGluIleIleIleTyrSerAspProGlyTyrHisSerGlnGluAspThr 120
 Db 301 GCATTTGATCTTGAATCATTTATTTACTCGACCTCGATATCATTTCAAGAAAGTACT 360
 QY 121 GluThrLysThrSerSerAsnGlyGluLysThrPheLeuTyrHisSerArgAlaIleThr 140
 Db 361 GAAACCAAGACATCTCGAATGAGAAAAAATTTCTTACCATTTCTGTCATTTACC 420
 QY 141 GluValGluLysThrGlyLeuGlySerSerAlaGlyLeuValSerValAlaIleThrSer 160
 Db 421 GAGGAGAAAGAACCCGATTTAGCTTATCGGACAGATTTAGTCTGCTTCCACACACT 480
 QY 161 LeuLeuSerHisPheIleProAsnValIleSerThrAsnLysAspIleLeuHisAsnVal 180
 Db 481 TTATATCCCATTTATATCCCAATGTATCATGATCAATTAAGATATTTGCAACAGTT 540
 QY 181 AlaGlnIleAlaHisCysTyrAlaGlnLysLysIleGlySerGlyPheAspValAlaThr 200
 Db 541 GCACAGATTTGACATTTGATATGCCCCAAAAAGAAAGATCTGGGTGTGATGTTCCAACT 600
 QY 201 AlaIleTyrGlyLeuIleValTyrArgArgPheGlnProAlaLeuIleAsnAspValPhe 220
 Db 601 GCATTTATGCTGATTTGATATGAAAGATTTTCAACCACTTTCATTAATTAACATGTTT 660
 QY 221 GlnValIleuGluSerAspProGluLysPheProThrGluLeuLysLeuIleGluSer 240
 Db 661 CAGGTTTAAAGAAAGATCGTGAAGAGTTCCCAAGAGTTGAAAATTTGATTCACAGT 720
 QY 241 AsnTrpGluGluLysHisGluArgCysThrLeuProTyrGlyIleLysLeuLeuMetGly 260
 Db 721 AACTGGGAATTCAAACATGAAAGATGTCATTACCAACGGAATCAAGTTATTAATGCGT 780
 QY 261 AspValLysGlyLysSerGluThrProLysLeuValSerArgValIleuGlnTrpLys 280
 Db 781 GACGTCAAGGGTGGCTCAAAACACCCAAATTTGATACAGATCTCAATGGAAGAAAG 840
 QY 281 GluLysProGluGluSerSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPhe 300
 Db 841 GAAAGCCAGAAAGAAAGCTGTTGTGATACCAAGCTTAATAGGCCAATTTACAGTTT 900
 QY 301 MetLysGluLeuAspGluMetArgGluLysTyrAspSerAspProGlnThrTyrIleLys 320
 Db 901 ATGAAGGAATGAGGAAATCGTGAAGAAATACAGACTCAGACCCAGACATTTATTTAA 960
 QY 321 GluLeuAspHisSerValGluProLeuThrValAlaIleLysAsnIleArgLysGlyLeu 340
 Db 961 GAGTTAGATCTTCTGTTGACCTTTGACTGTTGCGATTAAGAACATCAAGAAAGGGTTA 1020
 QY 341 GlnAlaLeuThrGlnLysSerGluValProIleGluProAspValGlnThrGlnLeuLeu 360
 Db 1021 CAAGCATTAACAAAAATACAGAGTTCCATTTGACCTGATGTCAAAACCAAGTTGTTG 1080
 QY 361 AsparGlySerGlnGluIleProGlyCysValGlyValValProGlyValGlyGlyTyr 380
 Db 1081 GACCGTTGTCAAGAGATTCCTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140

QY 381 AspAlaIleAlaValLeuValLeuGluAenGluValGlyAsnPhenylsThrLeu 400
DB 1141 GATGCAATAGCTGATTAAGTGTGGAAATCAAGTGGAAATTTTACGAGAACTCTT 1200
QY 401 GluAsnProAspTyrPheHisAenValTyrTrpValAspLeuGluGlnThrGluGly 420
DB 1201 GAAATCCAGATATATTTCTATATGTTTACTGGGTTGATTGGAGACAAACAGAGGT 1260
QY 421 ValLeuGluGlnLysProGluAspTyrIleGlyLeu 432
DB 1261 GTACTTGAAGAAACACAGACGATATATAGGTTTA 1296

RESULT 5

AAD02788/c
ID AAD02788 standard; DNA; 577 BP.

XX AAD02788;
XX
XX
XX 31-MAY-2001 (first entry)
XX
XX
XX S. cerevisiae ERG8 gene homologous DNA #2 from Candida albicans.
XX
XX Phosphomevalonate kinase; PMK; ERG8; anti-fungal agent; diagnosis;
XX infection; ds.
XX
XX Candida albicans.
XX
XX WO200114533-A2.
XX
XX 01-MAR-2001.
XX
XX 15-AUG-2000; 2000WO-GB03100.
XX
XX 21-AUG-1999; 99GB-0019766.
XX
XX (ASTR) ASTRAZENECA AB.
XX (ASTR) ASTRAZENECA UK LTD.
XX
XX Rosamond JDC, Schnell NF;
XX
XX WPI; 2001-218441/22.
XX
XX
XX New polypeptides and polynucleotides (ERG8) from Candida albicans,
XX PT useful in assays for identifying inhibitors of phosphomevalonate kinase
XX PT activity and as reagents for diagnosing C. albicans infection -
XX
XX
XX Claim 6; Page 25; 29pp; English.
XX
XX The patent discloses phosphomevalonate kinase (PMK; ERG8) protein
XX CC and their corresponding DNAs from Candida albicans. The ERG8 protein
XX CC is useful in assays for identifying compounds that inhibit phospho-
XX CC mevalonate kinase (PMK) activity. These inhibitors are useful as
XX CC anti-fungal agents. The ERG8 DNA and protein are also useful as
XX CC reagents for diagnosing C. albicans infection.
XX CC The present sequence is a Candida albicans DNA which is homologous
XX CC to the Saccharomyces cerevisiae ERG8 gene.
XX
XX
XX Sequence 577 BP; 167 A; 109 C; 108 G; 190 T; 3 other;

Alignment Scores:

Pred. No.: 2,96e-81 Length: 577
Score: 877.00 Matches: 182
Percent Similarity: 94.82% Conservative: 1
Best Local Similarity: 94.30% Mismatches: 7
Query Match: 39.33% Indels: 4
DB: 22 Gaps: 0

US-10-069-062-7 (1-432) x AAD02788 (1-577)

QY 61 GlyGluTrpGluTyrHisIleSerSerAenThrGluLysProArgGluValGlnSerArg 80
DB 573 GGGGAATGGGANTATCATATCATCATACNGA-AAACCCAAAGAGTNCAGTCACGC 515

QY 81 IleAsn-ProPheLeuGluAlaThrIlePheIle--ValLeuAlaTyrIleGlnProThr 99
DB 514 ATAAATCCCATTTTATAGAGCAACATATATCCATCGCTTTTATAGCTATATATCAACGACC 455
QY 100 GluAlaPheAspLeuGluIleIleIleTyrSerAspProGlyTyrHisSerGlnLysAsp 119
DB 454 GAAGCATTTGATCTTGAATCATTTATTTACTCGGACCCCTGGATATCATTCACAGAGAT 395
QY 120 ThrGluThrLysThrSerSerAsnGlyGluLysThrPheLeuTyrHisSerArgAlaIle 139
DB 394 ACTGAACACCAACATCTCTCGATGAGAGAGAGAAATCTTTCTTACCATTCCTGTCATT 335
QY 140 ThrGluValGluLysThrGlyLeuGlySerSerAlaGlyLeuValSerValAlaThr 159
DB 334 ACCGAAGTGAGAAACACCGGATTAGCTTCATCGCAGAGATTAGTGCAGTTGTTGCCACA 275
QY 160 SerLeuLeuSerHisPheIleProAsnValIleSerThrAsnLysAspIleLeuHisAsn 179
DB 274 AGTTATTTATCCCATTTTATCCCAATGTTTACAGTACGAATTAAGATATTTTGACAAAC 215
QY 180 ValAlaGlnIleAlaHisCysTyrAlaGlnLysLysIleGlySerGlyPheAspValAla 199
DB 214 GTTGCAACGATTCACATTTGTTATGCCAAGAAAGATAGAGATCTGCGTTGATGTGCA 155
QY 200 ThrAlaIleTyrGlyLeuIleValTyrArgArgPheGlnProAlaLeuIleAsnAspVal 219
DB 154 ACTGCAATTTATNGCTGATTTGATATATAGATTTTACGCCAGCTTGATTAATACCGTG 95
QY 220 PheGlnValLeuGluSerAspProGluLysPheProThrGluLeuLysLeuIleGlu 239
DB 94 TTTTCAGGTTCTGAGAAAGATCTCTGAGAAAGTTCCCAACAGAGTTGAAAAATGTGATGCA 35
QY 240 SerAenTrpGluLysHisGluLysCysThr 250
DB 34 AGTAACGTGGAAATTCAAACATGAAAGATGTACA 2

RESULT 6

AAD02787
ID AAD02787 standard; DNA; 547 BP.

XX AAD02787;
XX
XX
XX 31-MAY-2001 (first entry)
XX
XX
XX S. cerevisiae ERG8 gene homologous DNA #1 from Candida albicans.
XX DE
XX Phosphomevalonate kinase; PMK; ERG8; anti-fungal agent; diagnosis;
XX infection; ds.
XX
XX Candida albicans.
XX
XX WO200114533-A2.
XX
XX 01-MAR-2001.
XX
XX 15-AUG-2000; 2000WO-GB03100.
XX
XX 21-AUG-1999; 99GB-0019766.
XX
XX (ASTR) ASTRAZENECA AB.
XX (ASTR) ASTRAZENECA UK LTD.
XX
XX Rosamond JDC, Schnell NF;
XX
XX WPI; 2001-218441/22.
XX
XX
XX New polypeptides and polynucleotides (ERG8) from Candida albicans,
XX PT useful in assays for identifying inhibitors of phosphomevalonate kinase
XX PT activity and as reagents for diagnosing C. albicans infection -
XX
XX
XX Claim 6; Page 24; 29pp; English.
XX
XX The patent discloses phosphomevalonate kinase (PMK; ERG8) protein

CC and their corresponding DNAs from *Candida albicans*. The ERG8 protein is useful in assays for identifying compounds that inhibit phosphomevalonate kinase (PMK) activity. These inhibitors are useful as anti-fungal agents. The ERG8 DNA and protein are also useful as reagents for diagnosing *C. albicans* infection.

CC The present sequence is a *Candida albicans* DNA which is homologous to the *Saccharomyces cerevisiae* ERG8 gene.

XX
SQ Sequence 547 BP; 184 A; 81 C; 123 G; 159 T; 0 other;

Alignment Scores:

Pred. No.:	2,73e-75	Length:	547
Score:	819.00	Matches:	156
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	36.73%	Indels:	0
DB:	22	Gaps:	0

US-10-069-062-7 (1-432) x AAD02787 (1-547)

```
QY 277 GINTPLVLYSGILVSPROGLUGIUSERSERVAIVATYRASPQINLEUANSERALA 296
DB 2 CAATGAGAAAAGGAAAGCCAGAGAAAGCTCTGTTGTATGACCGATTATAGTCC 61
QY 257 AANLEUGINPHEMETLYSGILVUARGIUMETARGIULVYTYRASPSEAPPROGIU 316
DB 62 AATTTCAGTTTATGAGAGAAATTTGAGGAAATCGCGTAAATATGACATCAGACCGAG 121
QY 317 THTYTYLLEVSGILVUASPHISERVAIGIUPROLEUTHVALAILELYSANTLE 336
DB 122 ACTTATATTAAGATGATGATCACTTCTGTGAGCCCTTGACCTGTCGATTAGAACATC 181
QY 337 ATGLVSGILVLEUGIALALEUTHRGILVYSERGIUVALPROILEGUPROASPVAGIN 356
DB 182 AGAAGAAAGGTTTCAACGATTAACCAAAAATCAGAGCTTCAATTGAACTGATGTCNA 241
QY 357 THRGINLEUASPARGYSGLINGIUILEPROGIUYSVAGILVYVAIVALPROGIY 376
DB 242 ACCCAGTTGTTGACCGTTGTCAGAGATTCCTGGTGTGTTGGCGGTGTTCCAGGT 301
QY 377 AIAAGLYGYTYRASPALALEVALLEUVALLEUGIUAERGINVALGIVASNPHELYS 396
DB 302 GCTGTGTGATACGATGCAATACCTGATTAAGTGTGAAATCAAGTGGAAATTTTAAG 361
QY 397 GILVYSTRLEUGIUAENPROKAPTYRPHETISANVALTYRTTRVALASPLENGIUGI 416
DB 362 CAGAAACTCTTGAATAATCCAGATTATTTTCAATAAGTTTACTGGGTGATTGGAAGG 421
QY 417 GINTRGINGIUYVAILLEUGIUGIULVSPROGIUASPTYRILEGLYLEU 432
DB 422 CAAACAGAGGTGTACTTGAGAAAAAACAGAAAGACTATATAGTTTA 469
RESULT 7
AAL40805
ID AAL40805 standard; DNA; 1356 BP.
AC AAL40805;
XX
XX 03-OCT-2002 (first entry)
DE Nucleic acid relating to the production of prenyl alcohol SRQ ID No 29.
KW Prenyl alcohol; mutated cell; squalene synthase gene; geometric isomer;
KW industrial synthesis; isoprenoid-terpenoid compound; ds.
OS Saccharomyces cerevisiae.
XX
XX WO200253747-A1.
XX
XX 11-JUL-2002.
XX
XX 20-DEC-2001; 2001WO-JP11215.
XX
```

PR 28-DEC-2000; 2000JP-0401701.
PR 28-DEC-2000; 2000JP-0403067.
PR 18-SEP-2001; 2001JP-0282976.

XX (TOYOTA) TOYOTA JIDOSHA KK.

XX Ohto C, Obata S;

XX WPI; 2002-548086/58.

PT Production of prenyl alcohols by culturing translationally-active
PT mutated cells with reduced squalene synthase gene to express less
transcriptional product -

XX Disclosure; Page 237-238; 262pp; Japanese.

CC The invention relates to a method for producing prenyl alcohol comprising
CC culturing mutated cells, having been mutated so as to reduce the amount
CC of transcriptional product of squalene synthase gene transcriptional
CC activity, and then collecting prenyl alcohol from the culture medium. The
CC method is for the production of prenyl alcohols, which is for use in
CC industrial synthesis of isoprenoid-terpenoid compounds particularly
CC physiologically-active prenyl alcohol geometric isomers. This
CC polynucleotide sequence represents a nucleic acid sequence relating to
CC the method for producing prenyl alcohol comprising culturing mutated
CC cells.

SQ Sequence 1356 BP; 427 A; 236 C; 304 G; 389 T; 0 other;

Alignment Scores:

Pred. No.:	9.52e-74	Length:	1356
Score:	809.50	Matches:	190
Percent Similarity:	57.70%	Conservative:	76
Best Local Similarity:	41.21%	Mismatches:	146
Query Match:	36.30%	Indels:	49
DB:	24	Gaps:	14

US-10-069-062-7 (1-432) x AAL40805 (1-1356)

```
QY 3 LYSALAPHESERALAPROGLIYVYENALPHELEUALAGIYGYTYRLEUVALLEUGIUPRO 22
DB 13 AGAGCTTCACAGTGGCCCGGAGAAAGCGTTACTAGCTGTGATVTTAGTTTATGATACA 72
QY 23 ILETYRASPALATYRVALTHTRALAEUSERSERARQWETHISALVAILLETHPROLYS 42
DB 73 AAATATGAGCATTTGTGATCGATTAATCGCAAGAAAGCATGCTAGCCCATCTTAC 132
QY 43 GILYTHSERILEUYSGLUSER-----ARGILELYSILESERSEPPROGINPHEALA 59
DB 133 GGT---TCATTGCAAGGCTCGATTAAGTTGAAGTGGCTGGAAGAAAGTAAACAATTMAA 189
QY 60 AANGLYGLUTPGLIUTYRTHISLESERSEANTRGLU---LYSPROARGIULVALGIN 78
DB 190 GATGGGAGTGGCTGACCATATTAAGTCTTAAAGTGCCTTCATTCCTGTTTCGATAGGC 249
QY 79 SERARGILEANPPROPELEUGIUALATHRIIPHEILEVALLEUALATYRILEGLINPRO 98
DB 250 GGATCTAAGAACCTTTTCATTAAGAAAGTATTCCTCAAGTATTTAGCTTATTAACCT 309
QY 99 THR-----GLUALPHEASPLEUGIUILEILEIETYSERASP 111
DB 310 AACATGACGACTACTGCAATAGAAACTGTGCTTATATGAT-----ATTTCTCGAT 363
QY 112 PROGLIYTRHISERGINIUAERTHGLUTHRYSTRHISERSEANGLYGLIUYSTR 131
DB 364 GATGCCCTACCACTTCTGAGAGGATAGCGTTACCA-----CATCGGCGACAGAGA 417
QY 132 PHELEUTYRTHISERAGALAIETHTRGUVALGILVYSTRHISERSEANGLYGLIUYSTR 151
DB 418 TTGAGTTTCTTCTTGCAAGATTTGAAGAGCTTCCAAACAGGAGGCTGCTCTCGGCA 477
QY 152 GILYUVALSERVALATATHTSERLEUUSERHISPHLE-----PROASN 168
DB 111
```

```

Db 478 GGTTCAGTACAGTTTAACTACAGCTTGGCTCTCTTTTGTATGACGACTGAAAT 537
QY 169 ValIleSerThrAsnIleuHisAsnValAlaGlnIleLeuHisCysTyrAla 188
Db 538 AATGTAGACAAATATATAGAGAGATTATTCATTAATTTAGCACTATCTCATTTGCAAGCT 597
QY 189 GlnIlySerIleGlySerGlyPheAspValAlaThralleTyrGlyLeuIleValTyr 208
Db 598 CAGCGTAAATATGGAAGCGGTTTGTATGTAGCGCGGCGACATGTGATCTATACATAT 657
QY 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValleuGluSerAspProGlu 228
Db 658 AGAAGATTCACACCCCATTAATCTCAATTTCCAGATATTC-----GGAGTGGCT 708
QY 229 LysPheProThrGluLeuIlySerIleuIle-----GluSerAsnProGluGluIlySerGlu 247
Db 709 ACTTACGCGCATTAACCTGGCGCATTTGGTTGATGAAAGACATCGAATATTACGATTAA 768
QY 248 ArgCysThrLeuProTyrGlyIleIlySerLeuIleuMetGlyAspValIlySerGlu 267
Db 769 AGTAACCATTTACCTTCGGGATTAACCTTATGATGCGGCGATTAAGATGGTTTACAA 828
QY 268 ThrProIySerValSerArgValIleuGlnTyrIlySerGluIlySerProGluIlySer 287
Db 829 ACAGTAAACTGGTCCAGAGGTAAGAAATTTGGTATGATTCGCATATGCCAGAAAGCTTG 888
QY 288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
Db 889 AAAATATATACAGAACTCGATCTGCAATTTCTAGATTTATGATGACATTTAAACAT 948
QY 305 ArgGluMetArgGluIlySerAspSerAspProGluThrTyrIleIlySerGluIlySer 324
Db 949 GATCGCTTACAGAGACTGATGACATTCAGGATTCAGATTTAGTCTCTTGAAGAG 1008
QY 325 Ser-----ValGlnProLeuThrValAlaIleLysAsn 335
Db 1009 AATGACTGTACTGTCAAAAGTATCTGAATACAGAAATGTAGAGATGCAAGTTGCCACA 1068
QY 336 IleArgIlySerGluIleuGlnIleuThrGlnIlySerGluValProIleGluProAspVal 355
Db 1069 ATTAGACGTTCTTTGAAAAATATACATAAGAAATCTGGTCCGATATCGAACTCCCGTA 1128
QY 356 GlnThrGlnLeuLeuAspArgCysGlnIlyIleProGlyCysValGlyIlyValPro 375
Db 1129 CAACTAGCTTATGTGATGATGCTCCAGACCTTAAAGAGTTCTTCTGCTTAATACCT 1188
QY 376 GlyAlaGlyIlyTyrAspAlaIleAlaValleuVal-----LeuGluAsn 390
Db 1189 GGTGCTGGTGTATATGACCGCATTTGATGATTAAGCAGATGTTGATCTTGAAGGCT 1248
QY 391 GlnValGlyAsnPhelysGlnIlySerThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
Db 1249 CAATCCGCTATATACAAAGAA-----TTTCTTAAGTTTCAA 1284
QY 411 TrpValAspLeuGluIlyGlnThrGluIlyValleuGluIlySerProGluAspTyr 429
Db 1285 TGGCTGATGTATACGAGCTGACTGGGCTTGAAGAAAGAAAGATTCGGAAACTTAT 1344
QY 430 Ile 430
Db 1345 CTT 1347

```

```

RESULT 8
ID ABK96801
ID ABK96801 standard, DNA; 1356 BP.
XX
XX ABK96801;
XX
XX 24-SRP-2002 (first entry)
XX
XX S. cerevisiae prenyl diphosphate synthase gene #20.
XX
XX Prenyl alcohol; prenyl diphosphate synthase; geranylgeraniol;
XX

```

```

KM hydroxymethylglutaryl-CoA reductase; farnesylgeraniol; gene; ss;
KM isopentenyl diphosphate delta-isomerase; mevalonate kinase;
KM mevalonate CoA acetyltransferase; isoprenoid-terpenoid compound.
OS Saccharomyces cerevisiae.
XX
XX MO200253746-A1.
XX
XX 11-JUL-2002.
XX
XX 20-DEC-2001; 2001WO-0P11214.
XX
XX 28-DEC-2000; 2000JP-0403067.
XX
XX (TOYT ) TOYOTA JIDOSHA KK.
XX
XX Ohto C, Odata S, Muramatsu M, Nishi K, Totouka K;
XX
XX WPI; 2002-537944/57.
XX
XX Production of prenyl alcohols by culturing a transformant transferred
XX with e.g. prenyl diphosphate synthase gene, for use in industrial
XX synthesis of e.g. physiologically-active isoprenoid-terpenoid compounds
XX
XX Example 17; Page 269-270; 335pp; Japanese.
XX
XX The invention relates to a process for producing a prenyl alcohol
XX comprising: (A) construction of a recombinant by transferring an
XX expression recombinant DNA or a DNA for genome integration into a host
XX CC which contains prenyl diphosphate synthase gene or its variant; and
XX CC (B) collecting product from the culture medium. Also described is: (1) a
XX method for producing prenyl alcohol in which the expression
XX recombinant DNA contains: (a) a hydroxymethylglutaryl-CoA reductase
XX gene or its variant; or (b) an isopentenyl diphosphate delta-isomerase
XX (a) constructing a recombinant by transferring an expression
XX recombinant DNA or a DNA for genome integration into a host which
XX contains hydroxymethylglutaryl-CoA reductase gene or its variant; and
XX (b) isolating the product; (3) a process for producing farnesylgeraniol
XX in which the expression recombinant DNA also contains an isopentenyl
XX diphosphate delta-isomerase gene, mevalonate CoA acetyltransferase gene,
XX hydroxymethylglutaryl-CoA synthase gene, mevalonate kinase gene, or
XX mevalonate diphosphate decarboxylase gene. The methods are used for
XX the production of prenyl alcohols, particularly for use in industrial
XX synthesis of isoprenoid-terpenoid compounds e.g. physiologically-active
XX prenyl alcohols including geometric isomers. ABK96801-ABK96897
XX represent prenyl diphosphate synthase genes and related PCR primers used
XX in the methods of the invention.
XX
XX Sequence 1356 BP; 427 A; 236 C; 304 G; 389 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 9,52e-74 Length: 1356
XX Score: 809.50 Matches: 190
XX Percent Similarity: 57.70% Conserved: 76
XX Best Local Similarity: 41.21% Mismatches: 146
XX Query Match: 36,30% Indels: 49
XX DB: 24 Gaps: 14
XX
XX US-10-069-062-7 (1-432) x ABK96801 (1-1356)
QY 3 LysAlaPheSerAlaProGlyIlyValAlaPheLeuAlaGlyIlyTyrLeuValleuGluPro 22
Db 13 AAGAGCTTCAGTGGCCCGAGGAAAGCGTTACTAGCTGGGAGATTAATTTAGTTAATAC 72
QY 23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProIy 42
Db 73 AATATGAGCATTTGTATGATGATTAATGCGAAGATGCAATGCTGAGCCATCTTAC 132
QY 43 GlyThrSerLeuIlyGluSer-----ArgIleLysIleSerSerProGlnPheAla 59
Db 133 GGT---TCATTCGAAGGCTGTGATTAAGTTGAAGTGGCGGTGAAAAGTAAACATTTAA 189

```

```

QY 60 AaenGIyGluTrpGluTrpHisIleSerSerAsnThrGlu---LysProArgGluValGln 78
DB 190 GATGGGAGATGGCTGATACATATTAAGTCTTAAAGGCGCTTCATCTCTTCGATAGGC 249
QY 79 SerArgIleAsnProPheLeuGluValThrIlePheIleValLeuAlaTrpIleGlnPro 98
DB 250 GGATCTAAGAACCCCTTCATGGAAGTATGCGCTAAGCTATTTGATCTTAAACCT 309
QY 99 Thr-----GluAlaPheAspLeuGluIleIleIleIleIleIleIleIleIleIleIle 111
DB 310 AACATGACGACTGCTGCAATGAAAGCTGCTGATTTGATGAT-----ATTTCCTCTGAT 363
QY 112 ProGlyTrpHisSerGlnGluAspThrGluTrpHisSerSerAsnGlyGluLysTrp 131
DB 364 GATGCTTACATCTTCAGAGAGATAGCGCTTACCGAA-----CATCGTGGCAACAGAGA 417
QY 132 PheLeuTrpHisSerArgAlaIleThrGluValGluLysTrpGlyLeuGlySerSerAla 151
DB 418 TTGAGTTTTCATTCGACAGAAATTCGAAGATTCGCCAAACAGAGGCTGGGCTCTCGCA 477
QY 152 GlyLeuValSerValAlaIleThrSerLeuSerHisPheIle-----ProAsn 168
DB 478 GGTTTAGTCAAGTTTAACTACAGCTTGGCTCTCTTTTGTATCGGACCTGGAAT 537
QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTrpAla 188
DB 538 AATGTAGACAAATATAGAGAACTATCTCAATTTATGACCAAGTCTCTCATTTGCAAGCT 597
QY 189 GlnLysLysIleGlySerGlyPheAspValAlaIleThrGlyLeuIleValIleValTrp 208
DB 538 CAGGGTAAATTTGGAAGCGGTTTATGATGAGCGCGGCGCATATGATGATCTACAGAT 657
QY 209 ArgArgPheGlnProAlaIleuIleAsnAspValPheGlnValLeuGluSerAspProGlu 228
DB 658 AGAAGATTCACACCGCATTAATCTTAATTTGCCAGATATT-----GGAAGTGT 708
QY 229 LysPheProThrGluLeuLysLysLeuIle---GluSerAsnTrpGluLysHisGlu 247
DB 709 ACTTACGGAGAAACTGGCGCATTTGTGTATGAAAGAACTGCAATTTACGATTAA 768
QY 248 ArgCysThrLeuProTrpGlyIleLysLeuLeuMetGlyAspValLysGlySerGlu 267
DB 769 AGTAAACATTTTACCTCGGATTAACCTTATGATGAGCGGATATTAAGATGTTACAGA 828
QY 268 ThrProLysLeuValSerArgValLeuGlnTrpLysGluLysProGluGluSerSer 287
DB 829 ACAGTAAACTGCTCCAGAAAGGTAAATAATGTTGATGATTCGATATGCCAATAAGCTTG 888
QY 288 ValValIleAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
DB 889 AAAATATATACGAAGCTCATGATGCAATTTTAATTTATGATGAGCATATCTAACTA 948
QY 305 ArgGluMetArgGluLysTrpSerAspProGluTrpTrpIleLysGluLeuAspHis 324
DB 949 GATCGCTTAACAGACATCATGATGATGACAGATCATGATTTGATGATCTTACAGAG 1008
QY 325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
DB 1009 AATGACTGACTGTCAAAAGTATCTGAAACACAGAAAGTTAGATGACAGTTCACACA 1068
QY 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspAla 355
DB 1069 ATTACACGCTCTTATGAAAAATTAATCAAGATCTGTGCGGATATGACACTCCGTA 1128
QY 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
DB 1129 CAATCTACTATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1188
QY 376 GlyAlaGlyGlyTrpAspAlaIleAlaValLeuVal-----LeuGluAsn 390
DB 1189 GGTGCTGGTGTATGACGCCATTCAGATGATCTAAGCAAGATGTTGATCTTAAAGGCT 1248

```

```

QY 391 GlnValGlyAsnPheLysGlnLysThrLeuGluAsnProAspTrpPheHisAsnValTrp 410
DB 1249 CAATCGCTATATGACAAAAGA-----TTTTCTAAGTTCA 1284
QY 411 TrpValAspLeuGluGluGlnThrGluGlyValLeuGluGluLys---ProGluAspTrp 429
DB 1285 TGGCTGATGTATTAACACAGCTGATGCTGCTGTATGAAAGAAAAGATCCGGAATCTAT 1344
QY 430 Ile 430
DB 1345 CTT 1347

```

RESULT 9

AAD31013
ID AAD31013 standard; DNA; 1356 BP.

AAD31013;
31-MAY-2002 (first entry)

DE Yeast phosphomevalonate kinase (ERG8) orf.

KM Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;
KM isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;
KM transgenic plant; yeast; phosphomevalonate kinase; ERG8; PMK; de.

OS Saccharomycetes cerevisiae.

FN WO200210398-A2.

PD 07-FEB-2002.

PF 31-JUL-2001; 2001WO-US24037.

FR 31-JUL-2000; 2000US-221703P.

PA (HAHN/) HAHN P M.

PA (KUEH/) KUEHNLE A R.

PI Hahn FM, Kuehnle AR;

DR WPI; 2002-217122/27.

PT Use of specific genes of mevalonate and isoprenoid biosynthetic
PT pathways, for providing a cell with herbicide or antibiotic resistance,
PT and for providing transformed cells having increased isoprenoid
PT production.

PS Disclosure; Page 116-117; 193pp; English.

CC The invention relates to the use of specific genes of the mevalonate
CC and isoprenoid biosynthetic pathways and inactive gene sites
CC (pseudogene). Genes of the invention are used to enhance biosynthesis
CC of isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP)
CC and isoprenoid pathway derived products in the plastids of transgenic
CC plants and microalgae, for producing herbicide or antibiotic resistant
CC transgenic plants and microalgae, for providing transformed cells with
CC increased isoprenoid production compared to non-transformed cells, and
CC for providing a cell with an inserted polynucleotide sequence encoding
CC one or more products of interest. The present sequence is yeast
CC phosphomevalonate kinase (PMK; ERG8) EC 2.7.4.2. encoding orf.

XX Sequence 1356 BP; 427 A; 236 C; 304 G; 389 T; 0 other;

Alignment Scores:

Pred. No.: 9.52e-74 Length: 1356
Score: 809.50 Matches: 190
Percent Similarity: 57.70% Conservative: 76
Best Local Similarity: 41.21% Mismatches: 146
Query Match: 36.30% Indels: 49
DB: 24 Gaps: 14

US-10-069-062-7 (1-432) x AAD31013 (1-1356)

QY 3 LysAlaPheSerAlaProGluLysAlaPheLeuAlaGlyLysTyrLeuValLeuGluPro 22
 Db 13 AAGACCTTACAGTCCCGGAGGAGGCTTACAGTACGAGGATATTAGATTAGATAC 72
 QY 23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys 42
 Db 73 AATATGAAGACATTTGTATCGGATATATCGGACAGATATCATGCTGTACCATCTTAC 132
 QY 43 GlyThrSerLeuLysGluSer-----ArgIleLysIleSerSerProGluPheAla 59
 Db 133 GGT---TCATTCGACAGGCTGTATAGTTGAAGTCCGTGTAAGGATTAACATTTAA 189
 QY 60 AanglyGluTropGluTyrHisIleSerSerArgThrGlu---LysProGluValGlu 78
 Db 190 GATGGGAGTGGCTGTACCATATATAGTCTTAAAGGCTTACCTCTGTTCCGATAGGC 249
 QY 79 SerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIleGluPro 98
 Db 250 GGATCTAAGACCTTTTCATGTGAAAGATTATCGACGATATTAGTACTTAACT 309
 QY 99 Thr-----GluAlaPheAspLeuGluIleIleIleTyrSerAsp 111
 Db 310 AACATGACGACTACTGCAATAGAACTTGTCTGTATTGAT-----ATTCTCTGTAT 363
 QY 112 ProGlyTyrHisSerGluGluAspThrGluThrLysThrSerSerAsnGlyLysTyrThr 131
 Db 364 GATGCTTACATTTCTAGAGAGATAGCGTTACCGAA-----CATGTCGACACAGAA 417
 QY 132 PheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAla 151
 Db 418 TTGAGTTTCATTCGACAGAAATTGAAGAGTTCCCAAAACAGGGCTGGGCTCTCGGCA 477
 QY 152 GlyLeuValSerValValAlaThrSerLeuLeuSerHisPheIle-----ProAsn 168
 Db 478 GGTTTAGTCACAGTTTAACTACAGCTTGGCTCTTTTGTATCGACCTGGAAT 537
 QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGluIleAlaHisCysTyrAla 188
 Db 538 AATGTGACAAATATATAGAGATGATATTCATATTGACACAACTCTCATTTGTCAAGCT 597
 QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
 Db 598 CAGGGTAAATTCGAGCGGGTTGATGAGCGCGGACGATGATGATCAATAT 657
 QY 209 ArgArgPheGluProAlaLeuIleAsnAspValPheGluValLeuGluSerAspProGlu 228
 Db 658 AGAAGATCCACCCCATTAATCTTAATTGCGCAGATAT-----GGAGTGTCT 708
 QY 229 LysPheProThrGluLeuLysLysLeuIle---GluSerAspThrGluGluLysHisGlu 247
 Db 709 ACTTACGCGAGTAACTGGCGCATTTGGTTGATGAGAAAGACGTGAATATTACGATTAA 768
 QY 248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlySerGlu 267
 Db 769 AGTAAACATTACCTTCGGGATTAATCTTATGATGCGGAGATATTAAGATGGTTACAGA 828
 QY 268 ThrProLysLeuValSerArgValLeuGluIleTyrLysGluLysProGluGluSerSer 287
 Db 829 ACAGTAAACTGGTCCAGAAAGTAAAAATTTGTTATGATTCGATATGCGCAAAACCTTG 888
 QY 288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGluPheMet-----LysGluLeu 304
 Db 889 AAAATATATACAGAACTCATCATGCAAAATTCAGTTTATGATGAGATCATCTAACTA 948
 QY 305 ArgGluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHis 324
 Db 949 GATCGCTTACACAGAGCTCATGACGATTAACAGATCAAGATATTGAGCTCTTGAGAG 1008
 QY 325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
 Db 1009 AATGACTGTACTGTCAAAAGTATCTGTAATCAAGAAATTAGATGATGATGATGACAGA 1068

QY 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspVal 355
 Db 1069 ATTACACCTTCTTTAGAAAAATTAACATAAGATTTGGTGGCATATGAACTCCGTA 1128
 QY 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
 Db 1129 CAATACCTTATATGATGATATGCGACCTTAAAGAGATTTCTTACTTGTATATACCT 1188
 QY 376 GlyAlaGlyLysTyrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
 Db 1189 GGTCTGTGTGTATAGACGCACTGCACTGATTAACAGCAAGATGTTGATAGGCT 1248
 QY 391 GlnValGlyAsnPheLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
 Db 1249 CAACCGCTAATGACAAAG-----TTTCTAAGCTTCA 1284
 QY 411 TrpValAspLeuGluGluGlnThrGluGlyValLeuGluGluLys---ProGluAspTyr 429
 Db 1285 TGGCTGTGTATTAACACAGCTGACTGGGTGTAGAAAGAAAGATCCGAAACTTAT 1344
 QY 430 Ile 430
 Db 1345 CTT 1347
 QY 430 Ile 430
 Db 1345 CTT 1347
 RESULT 10
 AAD31026
 ID AAD31026 standard; DNA; 7681 BP.
 AC AAD31026;
 DT 31-MAY-2002 (first entry)
 DE Operon D DNA encoding mevalonate pathway.
 KW Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;
 KW isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;
 KW transgenic plant; yeast; phosphomevalonate kinase; HMGRT; HMGs; AACT;
 KW mevalonate kinase; mevalonate diphosphate decarboxylase; HMG-CoA; MD;
 KW acetoacetyl thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;
 KW MVK; PKM; HMG-CoA reductase; ds.
 OS Chimeric - Saccharomyces cerevisiae.
 OS Chimeric - Arabidopsis thaliana.
 OS Chimeric - Streptomyces sp.
 PN W0200210398-A2.
 PD 07-FEB-2002.
 PF 31-JUL-2001; 2001WO-US24037.
 PR 31-JUL-2000; 2000US-221703P.
 XX (HAHN/) HAHN F M.
 XX (KUEH/) KUEHNLE A R.
 XX Hahn FM, Kuehnle AR;
 DR WPI, 2002-217122/27.
 PT Use of specific genes of mevalonate and isoprenoid biosynthetic
 PT pathways, for providing a cell with herbicide or antibiotic resistance,
 PT and for providing transformed cells having increased isoprenoid
 PT production -
 PS Claim 77; Page 141-145; 193pp; English.
 CC The invention relates to the use of specific genes of the mevalonate
 CC and isoprenoid biosynthetic pathways and inactive gene sites
 CC (pseudogene). Genes of the invention are used to enhance biosynthesis
 CC of isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP)
 CC and isoprenoid pathway derived products in the plastids of transgenic
 CC plants and microalgae, for producing herbicide or antibiotic resistant

transgenic plants and microalgae, for providing transformed cells with increased isoprenoid production compared to non-transformed cells, and for providing a cell with an inserted polynucleotide sequence encoding one or more products of interest. The present sequence is operon D. DNA encoding the entire mevalonate pathway. This operon contains S. cerevisiae orf encoding phosphomevalonate kinase (PMK), mevalonate kinase (MVK), mevalonate diphosphate decarboxylase (MDD), acetosuccinyl chiolase (AACT) and A. thaliana orf encoding 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) synthase (HMGs) and Streptomyces sp. CL190 orf encoding HMG-CoA reductase (HMGrt).

XX Sequence 7681 BP; 2127 A; 1691 C; 1837 G; 2026 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	7681
Score:	Matches:	190
Percent Similarity:	Conservative:	76
Best Local Similarity:	Mismatches:	146
Query Match:	Indels:	49
	Gaps:	14

US-10-069-062-7 (1-432) x AAD31026 (1-7681)

```

Oy 3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyGlyTyrLeuValLeuGluPro 22
Db 47 AGAGCCTTCAGTCCCGCCAGGAAAGCGTACTAGCTGGATATTAGTTTATGATACA 106
Oy 23 LLeTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProlys 42
Db 107 AAATGAAAGCATGTGTAGTTCGATATTCGCAAGAAATCATGCTGATGCCCATCTTAC 166
Oy 43 GlyThrSerLeuLysGluSer-----ArgIleLysIleSerSerProGluPheAla 59
Db 167 GGT---TATTGCAAGGGTCTGATAGTTTGAAGCGGTGAAAGAAATTAACATTTAA 223
Oy 60 AenGlyGluTyrGlyLysIleSerSerAsnThrGlu---LysProArgGluValGln 78
Db 224 GATGGGAGGTGCTTACCATTAAGTCTTAAAGTGGCTTCATTCCTGTTTCGATAGGC 283
Oy 79 SerArgIleAsnProPheLeuGluLysIleThrIlePheIleValLeuAlaTyrIleGlnPro 98
Db 284 GGATCTAAGAACCTTTTCATTGAAAAGTTATCGCTTAACGTAATTTAGCTTTAAACCT 343
Oy 99 Thr-----GluAlaPheAspLeuGluIleIleIleTyrSerAsp 111
Db 344 AACATGAGCAGTACTGCATAGAAACTGTTCTGTTATGAT-----ATTTCCTGAT 397
Oy 112 ProGlyTyrHisSerGlnGluAspThrGluThrLysThrSerSerAsnGlyGluLysThr 131
Db 398 GATGCTTACCATTTCTCAGGAGGATAGCGTTACCGAA-----CATCGTGGCAAGAGA 451
Oy 132 PheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAla 151
Db 452 TTGAGTTTTCATTCGACAGAAATTAAGAAAGTTCCCAAAACGGCTGGGCTCCCGCA 511
Oy 152 GlyLeuValSerValAlaIleThrSerLeuSerSerHisPheIle-----ProAsn 168
Db 512 GGTTCATGACAGTTTACAGCTTTGGCTCTTTTGTATTCGACCTGGAAT 571
Oy 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
Db 572 AATGATGACAAATATAGAGAGAGTATTCATATTTAGACAAAGTTGCTCATGTCAACT 631
Oy 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
Db 632 CAGGGTAAATATGAGAGCGGTTTATGATGAGCGGCGCATATGATCTATGATAT 691
Oy 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGlu 228
Db 692 AGAAGATTCGCCCGCATTAATCTTAATTTGCCAGATATT-----GGAAGTCT 742
Oy 229 LysPheProThrGluLeuLysLysLeuIle---GluSerAsnTyrGluGluLysHisGlu 247
Db 247 LysPheProThrGluLeuLysLysLeuIle---GluSerAsnTyrGluGluLysHisGlu 247

```

```

Db 743 ACTTACGGCAGTAAATCGCCGCAATTGGTTGATGAGAAGACTGGAATATTACGATTTAA 802
Oy 248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlyValSerGlu 267
Db 803 ACTAACCATTTTACCTTCGGGATTAATCTTTATGATGGCGGATATTAAAGATTCACAGA 862
Oy 268 ThrProLysLeuValSerArgValLeuGlnIleTyrLysLysGlyLysProGluLysSer 287
Db 863 ACAGTAAATCGTTCAGAGAGTAAAGTAAATTTGATGATTCGATATCCAGAAAGCTTG 922
Oy 288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
Db 923 AAATATATATACAACTGATCATGCAATTCGCAAAATTCGATTTATGATGATCTTAACCT 982
Oy 305 ArgGluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHis 324
Db 983 GATGCTTACACAGACTCATGACATTCACAGCATCAGATTCGATTTGATCTCTTGAGAG 1042
Oy 325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
Db 1043 AATGACGTACCTGCAAAAGATTCCTGAATTCACAGAAAGTTAGATGACAGTGGCCACA 1102
Oy 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspVal 355
Db 1103 ATTAGACGTTCTTTAGAAATATTAATCAATCTGATGCGGATATGAACTCCCGTA 1162
Oy 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
Db 1163 CAATCTACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1222
Oy 376 GlyAlaGlyGlyTyrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
Db 1223 GGTGCTGTGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1282
Oy 391 GlnValGlyAsnPheLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
Db 1283 CAACCGCTTAATGACAAAGA-----TTTTCTAAGGTTCA 1318
Oy 411 TrpValAspLeuGluGlnGluThrGluGlyValLeuGluGluLys---ProGluAspTyr 429
Db 1319 TGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1378
Oy 430 Ile 430
Db 1379 CTT 1381

RESULT 11
AAD31023 standard; DNA; 7693 BP.
AAD31023;
31-MAY-2002 (first entry)

Operon A DNA encoding mevalonate pathway.

Mevlonate; isopentenyl diphosphate; hebbicide resistance; DMAPP; IPP;
isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;
transgenic plant; yeast; phosphomevalonate kinase; HMGs; AACT;
mevalonate kinase; mevalonate diphosphate decarboxylase; MDD;
acetosuccinyl thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;
MVK; PMK; HMG-CoA reductase; ds.

OS Chimeric - Saccharomyces cerevisiae.
OS Chimeric - Arabidopsis thaliana.
PN MO200210398-A2.
PD 07-FEB-2002.
PF 31-JUL-2001; 2001WO-US24037.
PR 31-JUL-2000; 2000US-221703P.

```

XX (HAHN/) HAHN F M.
 PA (KUEHNLE) KUEHNLE A R.
 XX Hahn FM, Kuehnle AR;
 XX WPI; 2002-217122/27.
 DR Use of specific genes of mevalonate and isoprenoid biosynthetic
 PT pathways, for providing a cell with herbicide or antibiotic resistance,
 PT and for providing transformed cells having increased isoprenoid
 PT production
 PS Claim 77; Page 127-131; 193pp; English.
 XX The invention relates to the use of specific genes of the mevalonate
 CC and isoprenoid biosynthetic pathways and inactive gene sites
 CC (pseudogene). Genes of the invention are used to enhance biosynthesis
 CC of isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP)
 CC and isoprenoid pathway derived products in the plastids of transgenic
 CC plants and microalgae, for producing herbicide or antibiotic resistant
 CC transgenic plants and microalgae, for providing transformed cells with
 CC increased isoprenoid production compared to non-transformed cells, and
 CC for providing a cell with an inserted polynucleotide sequence encoding
 CC one or more products of interest. The present sequence is operon A DNA
 CC encoding the entire mevalonate pathway. This operon contains
 CC S. cerevisiae orf encoding phosphomevalonate kinase (PMK), mevalonate
 CC kinase (MVK), mevalonate diphosphate decarboxylase (MDD), acetoacetyl
 CC chiolzyme (AACT) and A. thaliana orf encoding 3-hydroxy-3-methylglutaryl-
 CC coenzyme A (HMG-CoA) synthase (HMGs) and HMG-CoA reductase (HMGRL).
 XX Sequence 7693 BP; 2212 A; 1526 C; 1794 G; 2161 T; 0 other;
 SQ

Alignment Scores:
 Pred. No.: 1,12e-72 Length: 7693
 Score: 809.50 Matches: 190
 Percent Similarity: 57.70% Conservative: 76
 Best Local Similarity: 41.21% Mismatches: 146
 Query Match: 36.30% Indels: 49
 DB: 24 Gaps: 14

US-10-069-062-7 (1-432) x AAD31023 (1-7693)

QY 3 LysAlaPheSerLeuAlaProGlyLysAlaPheLeuAlaGlyLysLeuValLeuGluPro 22
 Db AGAGCTTACAGTCCCGGAGGAAAGGCTTACTAGCTGCGATTTAGTTAGTTAGTACA 3859
 QY 23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys 42
 Db AATATGAAGCATTTGTAGTCGATTCGCAAGATGATGCTAGCCCATCTTAC 3919
 QY 43 GlyThrSerLeuLysGluSer-----ArgIleLysIleSerSerProGlnPheAla 59
 Db 3920 GGT---TCATTTGCAAGGGTGTGTAAGTTGAAGTGGTGTGAAAGTAAACATTATAA 3976
 QY 60 AsnGlyValTyrPgluTyrHisIleSerSerSerSerSerGlu---LysProArgGluValGln 78
 Db 3977 GATGGGAGAGTGGCTGTAACCTAATAAGTCTTAAGTGGCTTCTCTTTCGATGAGCC 4036
 QY 79 SerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnPro 98
 Db 4037 GCATCTAAGAACCTTTTCATTTGAAAGATATGCTTACGATTTAGCTTAACCTTAAACCT 4096
 QY 99 Thr-----GluAlaPheAspLeuGluIleIleIleTyrSerAsp 111
 Db 4097 AACATGACGACTACTGCAATAGATAAATTGTTGTTATTCAT---ATTTCCTCTGAT 4150
 QY 112 ProGlyTyrHisSerGlnGluAspThrGluThrLysThrSerSerAsnGlyGluLysThr 131
 Db 4151 GATGCTACCATTTCTGAGGAGATAGCTTACCGAA-----CATGTGGCAACAGAGA 4204
 QY 132 PheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAla 151
 Db 151

Db 4205 TTGAGTTTTCATTCGACAGATTTGAGAGAGTCCCAAGAGGCGCTCCTCGGCA 4264
 QY 152 GlyLeuValSerValValAlaThrSerLeuLeuSerHisPheIle-----ProAsn 168
 Db 4265 GGTTAGTACACAGTTTAACTTAACAGCTTTGGCTCTTTTGTATTCGACCTGAAAT 4324
 QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
 Db 4325 AATGTAGCAAAATATAGAGAGATTTATTCATATTTAGCACAGATGCTCATGTCAACT 4384
 QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
 Db 4385 CAGGGTAAATTTGGAAAGGGGTTGATGTAGCGGCGAGCATATGATCTATCAGATAT 4444
 QY 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGluValLeuGluSerAspProGlu 228
 Db 4445 ACAAGATTCACACCCGCTTATTTCTTAATTTCCAGATATTT-----GGAAGTCT 4495
 QY 229 LysPheProThrGluLeuLysLysLeuIle---GluSerAsnTyrPgluGluLysIleGlu 247
 Db 4496 ACTTAGCGGAGTAACCTGGCGGATTTGGTGTAGTGAAGAGACTGGAATTTAGCATTA 4555
 QY 248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlySerGlu 267
 Db 4556 AGTAACCATTTACTTCGCGATTTAATTATGATGAGCGGATATTTAGAAATGCTTACAA 4615
 QY 268 ThrProLysLeuValSerArgValLeuGlnIleTyrLysGluLysProGluGluSerSer 287
 Db 4616 AAGATTAACCTGCTGACAGAGTTAAATAATTTGTATGTGATTCGATATGCCAGAAAGCTTG 4675
 QY 288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
 Db 4676 AATATATATACAGAACTGATCATCAATCTAATTTATGATGATGATGATCTAATCACTA 4735
 QY 305 ArgGluMetArgGluLysTyrAspSerAspProGluTyrIleLysGluLeuAspHis 324
 Db 4736 GATGCTTACACAGACTCATGATGATTAACAGCGATTCAGATATTTGAGTCTTTAGAGG 4795
 QY 325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
 Db 4796 AATGACTGATCTGCTCAAAAGTATCTGAAATCACAGAAAGTTAGATGACGATGCCACA 4855
 QY 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspVal 355
 Db 4856 ATTAGACGTTCTCTTAAAGAAATTAATAAGATCTGTCGCGATATGAACTCCCGTA 4915
 QY 356 GlnThrGlnLeuLeuAspArgCysGlnGlnIleLeuProGlyCysValGlyValValPro 375
 Db 4916 CAATCTACTTAATTTGGATGATGATCCAGACTTTAAAGAGATTTCTTAATTAATCT 4975
 QY 376 GlyAlaGlyGlyTyrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
 Db 4976 GGTGCTGATGTTATGAGCGCATTTCCAGTATTTACTAAGCAAGATGTTATGAGGCT 5035
 QY 391 GlnValGlyAsnPheLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
 Db 5036 CAACCGCTAATAGACAAAGA-----TTTTCTAAGTTCA 5071
 QY 411 TrpValAspLeuGluGlnGlnThrGluLysValLeuGluGluLys---ProGluAspTyr 429
 Db 5072 TGGCTGATGTAACTCAGCTGATCTGGGTTTAAAGAAAGAAAGATCCGAAACTTAT 5131
 QY 430 Ile 430
 Db 5132 CTT 5134
 Db 5132 CTT 5134

RESULT 12
 AAD31024
 ID AAD31024 standard; DNA; 7695 BP.
 AC AAD31024;
 XX
 DT 31-MAY-2002 (first entry)

QY 391 GlnValGlyAsnPhenylGlnThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
DB 1268 CAAACCGCTTAATACAAAGAA-----TTTCTAAGTTTCA 1303
QY 411 TTPValAspLeuGlnGluGlnThrGluGlyValleuGluGlySerProGluAspTyr 429
DB 1304 TGGCTGATGTAATCACTGAGCTGCTGCTGTTAGAGAAAGAAATCGGAACTTAT 1363
QY 430 Ile 430
DB 1364 CTT 1366
RESULT 13
AAD31027
ID AAD31027 standard; DNA: 8224 BP.
AC AAD31027;
XX
XX 31-MAY-2002 (first entry)
DE Operon B DNA encoding mevalonate pathway and IPP isomerase.
XX Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;
XX isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;
XX transgenic plant; Yeast; phosphomevalonate kinase; HMGb; HMGs; AACT;
XX mevalonate kinase; mevalonate diphosphate decarboxylase; HMG-CoA; MDD;
XX acetoacetyl thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;
XX MVK; PMK; HMG-CoA reductase; IPP isomerase; IPPi; ds.
XX
XX Chimeric - Saccharomyces cerevisiae.
OS Chimeric - Arabidopsis thaliana.
OS Chimeric - Streptomyces sp.
OS Chimeric - Rhodospirillum rubrum.
XX W0200210398-A2.
XX
XX 07-FEB-2002.
XX
XX 31-JUL-2001; 2001MO-US24037.
XX
XX 31-JUL-2000; 2000US-221703P.
XX
XX (HAHN/) HAHN F M.
XX (KUEHNLE) KUEHNLE A R.
XX
XX Hahn FM, Kuehnle AR;
XX
XX WPI; 2002-217122/27.
XX
XX Use of specific genes of mevalonate and isoprenoid biosynthetic
XX pathways, for providing a cell with herbicide or antibiotic resistance,
XX and for providing transformed cells having increased isoprenoid
XX production -
XX
XX Claim 77; Page 145-149; 193pp; English.
XX
XX The invention relates to the use of specific genes of the mevalonate
XX and isoprenoid biosynthetic pathways and inactive gene sites
XX (pseudogenes). Genes of the invention are used to enhance biosynthesis
XX of isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP)
XX and isoprenoid pathway derived products in the plasmids of transgenic
XX plants and microalgae, for producing herbicide or antibiotic resistant
XX transgenic plants and microalgae, for providing transformed cells with
XX increased isoprenoid production compared to non-transformed cells, and
XX for providing a cell with an inserted polynucleotide sequence encoding
XX one or more products of interest. The present sequence is operon B DNA
XX encoding the entire mevalonate pathway. This operon contains
XX S. cerevisiae orf encoding phosphomevalonate kinase (PMK), mevalonate
XX kinase (MVK), mevalonate diphosphate decarboxylase (MDD), acetoacetyl
XX thiolase (AACT), and A. thaliana orf encoding 3-hydroxy-3-methylglutaryl-
XX coenzyme A (HMG-CoA) synthase (HMGs); Streptomyces sp. Cl.90 orf
XX encoding HMG-CoA reductase (HMGb) and R. capsulatus DNA encoding IPP

CC isomerase (IPPI).
XX
SQ Sequence 8224 BP; 2228 A; 1847 C; 2016 G; 2133 T; 0 other;
Alignment Scores:
Pred. No.: 1,236-72 Length: 8224
Score: 809.50 Matches: 190
Percent Similarity: 57.70% Conservative: 76
Best Local Similarity: 41.21% Mismatches: 146
Query Match: 36.30% Indels: 49
DB: 24 Gaps: 14
US-10-069-062-7 (1-432) x AAD31027 (1-8224)
QY 3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyTyrLeuValleuGluPro 22
DB 47 AGAGCTTCAAGTCCCGGAGGAAACGTTACTAGCTGATATTTAGTTAATACATA 106
QY 23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProlys 42
DB 107 AATATGAAGCATTTGTAGTCGATTAATCGGCAAGATGATGCTGAGCCATCTTAC 166
QY 43 GlyThrSerLeuLysGluSer-----ArgIleLysIleSerSerProGlnPheAla 59
DB 167 GGT---TCATTCGCAAGGCTGTGATTAAGTGAAGTGCCTGCAAAAGTAAACATTTAA 223
QY 60 AanglyGluTyrGluTyrHisIleSerSerSerSerSerSerSerSerSerSerSer 78
DB 224 GATGAGGAGTGTCTTACATTAAGTAAAGTCTTAAGTCTTATTCGTTTTCATGAGGC 283
QY 79 SerArgIleAsnProPheLeuGluAlaThrIlePheIleValleuAlaTyrIleGlnPro 98
DB 284 GGATCTAGAAACCTTTTATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 343
QY 99 Thr-----GluAlaPheAspLeuGluIleIleTyrSerAsp 111
DB 344 AACATGACAGTACTGCAATTAAGTAAAGTCTTATTCGTTATTCGAT-----ATTTCTCTGAT 397
QY 112 ProGlyTyrHisSerSerGluAspThrGluThrIleSerSerSerSerSerSerSer 131
DB 398 GATGCTTCAAGTCCCGGAGGAAACGTTACTAGCTGATATTTAGTTAATACATA 451
QY 132 PheLeuTyrHisSerArgAlaIleThrGluValGluTyrGluGlySerSerAla 151
DB 452 TTGAGTTTCAATTCGACAGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 511
QY 152 GlyLeuValSerValAlaIleThrSerLeuSerHisPheIle-----ProAsn 168
DB 512 GGTTCATGTCAGATTTTAACTACAGCTTTGGCTCTTTTGTATTCGACCTGAAAT 571
QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
DB 572 AATGTAAGCAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 631
QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
DB 632 CAGGTAATTAATGGAAGCGGTTTATGATGAGGCGGACATATGATATTCATGATCA 691
QY 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValleuGluSerAspProGlu 228
DB 692 AAAAAATTCACCGCATTAATCTTAATTCGACAGATTT-----GGAAGTGTCT 742
QY 222 LysPheProThrGluLeuLysIleLeu-----GluSerAsnTyrGluLysHisGlu 247
DB 743 ACTTACGGGAGTAACTGCGCATTTGTATGTAAGAAAGTGAATTAATTAATTAATTA 802
QY 248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetCysIleAspValIleGlySerGlu 267
DB 803 ACTAATCAATTAATCTTCGATTAATTCATTAATTCATTAATTCATTAATTCATTAAT 862
QY 268 ThrProLysLeuValSerArgValleuGlnTyrLysGluLysProGluGluSerSer 287
DB 863 ACAGTAAATGCTGTCAGAGGTAATAAATGATGATGATTCGATATTCGATATTCGATATTCG 922


```

Db      557 AATGTGACAAATATAGAGAAATTATTCATTAATTGACCAAGTGTCTCATTTCAAGCT 616
      189 Glnlylsyllelglyserglypneaspvalalethralalelyrglyleuilevaltyr 208
      617 CAGGCTAAATTCGAAAGCGGCTTGTGATGACGCGCATATGATCATCTACATCATAT 676
      209 ATGATGPhelInProAlaLeuileasnaspvalPheGlnValLeuIngluserAProGlu 228
      677 AGAAGATTCACCCACCCGATTAATCTTAATTCGAGATAT-----GGAAGTCT 727
      229 LysPheProThrgluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 247
      728 ACTTACGCGAGTAAGTACGCGCATTTGGTTGATGAAAGAAAGACTGGAATATACAGATTAA 767
      248 ArgCysThreuleuProtyrGlylylelyleuLeuMetGlyAspValysglysergly 267
      788 AGTACCATTTTACCTTCGGATTAACCTTTATGATGATGATGATGATGATGATGATGAT 847
      268 ThrProlyseuleuValserArgValleuGlnTrpLysleuGlyserglyserglyser 287
      848 ACGTAAACTGTCGCAAGAGTAAATAATGATGATGATGATGATGATGATGATGATGATG 907
      288 ValValTyraSpInleuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 304
      908 AATATATATACAGAACTTCATCATGCAAAATTCATGATTAATGATGATGATGATGATG 967
      305 ArgGluMetArgGlyserTyraSpSerAProGluThrTyrllelyserglyleuLeuA 324
      968 GATCGCTTACACAGACATCATGATGATGATGATGATGATGATGATGATGATGATGATG 1027
      325 Ser-----ValGluProleuThrValAlaIleLysAsn 335
      1028 AATGACTGATACCTGCAAAAGTATCCTGAAATACACAGAACTTGAAGATGATGATG 1087
      336 IleArglyserglyleuGlnAlaLeuThrGlnLysSerGlyValProIleGluProA 355
      1088 ATTAGAGCTTCCTTGAATAAATAATTAAGAAATGATGATGATGATGATGATGATGATG 1147
      356 GlnThrGlnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 375
      1148 CAACATGATCTTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1207
      376 GlyAlaGlyValTyraSpAlaIleAlaValLeuVal-----LeuGluAsn 390
      1208 GGTGCTGGTGTATATGACCCATTCGATGATGATGATGATGATGATGATGATGATGATG 1267
      391 GlnValGlyAsnPhelysGlnLysThrLeuGluAsnProAAspTyraPheHisAsnVal 410
      1268 CAACCGCTAATGACAAAGA-----TTTCTAAGTTCA 1303
      411 TrpValAspLeuGlnGlnGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 429
      1304 TGGCTGATGTACTACAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1363
      430 Ile 430
      1364 CTT 1366

```

```

OS      Chimeric - Saccharomyces cerevisiae.
OS      Chimeric - Arabidopsis thaliana.
OS      Chimeric - Schizosaccharomyces pombe.
XX      WO200210398-A2.
XX      07-FEB-2002.
XX      31-JUL-2001; 2001WO-US24037.
XX      31-JUL-2000; 2000US-221703P.
XX      (HAIN/) HAIN F M.
XX      (KUEH/) KUEHNLE A R.
XX      Hain FM, Kuehnle AR;
XX      WPI; 2002-217122/27.
XX      Use of specific genes of mevalonate and isoprenoid biosynthetic
XX      pathways, for providing a cell with herbicide or antibiotic resistance,
XX      and for providing transformed cells having increased isoprenoid
XX      production.
XX      Claim 77; Page 154-158; 193pp; English.
XX      The invention relates to the use of specific genes of the mevalonate
XX      and isoprenoid biosynthetic pathways and inactive gene sites
XX      (pseudogene). Genes of the invention are used to enhance biosynthesis
XX      of isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP)
XX      and isoprenoid pathway derived products in the plastids of transgenic
XX      plants and microalgae, for producing herbicide or antibiotic resistant
XX      transgenic plants and microalgae, for providing transformed cells with
XX      increased isoprenoid production compared to non-transformed cells, and
XX      for providing a cell with an inserted polynucleotide sequence encoding
XX      one or more products of interest. The present sequence is operon G
XX      encoding entire mevalonate pathway and ID11 (IPP isomerase). This
XX      operon contains S. cerevisiae, A. thaliana and S. pombe DNA.
XX      SO      Sequence 8400 BP; 2437 A; 1644 C; 1941 G; 2378 T; 0 other:
XX      Alignment Scores:
XX      Pred. No.: 1,276-72 Length: 8400
XX      Score: 809.50 Matches: 190
XX      Percent Similarity: 57.70% Conservative: 76
XX      Best Local Similarity: 41.21% Mismatches: 146
XX      Query Match: 36.30% Indels: 49
XX      DB: 24 Gaps: 14
XX      US-10-069-062-7 (1-432) x AAD31029 (1-8400)
      QY      3 LysAlaPheSerAlaProGlyValPheLeuAlaGlyValTyrlleuValleuGluPro 22
      DB      32 AAGACCTTATAGTCCCGGAGAAAGCGTATCTAGCTGGAGATATTTAGTTAGATACA 91
      QY      23 IleTyraSpAlaTyraValThrAlaLeuSerSerArgMetHisAlaValIleThrProlys 42
      DB      92 AATATGAAAGCATTTGTAGTCGATTAATCGCAAGAAATGCAATGATGCCATCTTAC 151
      QY      43 GlyThrSerLeuLeuGluSer-----ArgIleLysIleSerSerProGlnPheAla 59
      DB      152 GGT---TCATTCGAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 208
      QY      60 AsnGlyGlnTrpGlnTyrllelelelelelelelelelelelelelelelelelelele 78
      DB      209 GATGGGAGTGGCTGATCAATTAAGTCTTAAGTGGCTTCAATTCCTTTCGATAGGC 268
      QY      79 SerArgIleAsnProPheLeuGlnAlaThrIlePheIleValleuAlaTyrlleGlnPro 98
      DB      269 GATCTACAGAACCTTTCTATTGAAAAGTATCGCTAAGCTATTTAGCTTAAACT 328
      QY      99 Thr-----GluAlaPheAspLeuGlnIlelelelelelelelelelelelelelelele 111

```

```

Db      329 AACATGACGACTACTGCATAGAAAATTGCTGTTATTGAT-----ATTTCCTCTGAT 382
Qy      112 ProGlyTyrHisSerGlnGluAspThrGlnTyrHisSerSerAsnGlyGluLysThr 131
Db      383 GATGCTACCATCTTCAGAGAGATAGCGTTACCGAA-----CANCCTGGCAACAGAGA 436
Qy      132 PheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAla 151
Db      437 TTGAGTTTCATCTCCACAGAAATGAGAAATGCCAAACAGGGCTGGCTCTCCGCGCA 496
Qy      152 GlyLeuValSerValValAlaIleThrSerLeuSerHisPheIle-----ProAsn 168
Db      497 GGTTCAGTCACAGTTTACCTACAGCTTGGCCCTCTTTTGTATCGGACCTGGAAAT 556
Qy      169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
Db      557 AATGAGCAATATATAGAAAGAACTTATCTCATTTAGCAACAGTTGCTCATTTGCAAGT 616
Qy      189 GlnLysLysIleGlySerGlyPheAspValAlaIleThrAlaIleTyrGlyLeuValTyr 208
Db      617 CAGGCTAAATTTGGAAGCGGGTTTGTATGACGCGCGGCGCATATGATCTATCAGATAT 676
Qy      209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGlnSerAspProGlu 228
Db      677 AGAAGATTCACCCCGCATTAATCTTAATTTGCCAGATAT-----GAAAGTGCT 727
Qy      229 LysPheProThrGlnLysLysLeuIle---GluSerAsnTrpGlnGluLysHisGlu 247
Db      728 ACTTACGCGCAGTAACTGGCCGATTTGGTGTATGAGAAAGAACTGAAATTTACGATTAA 787
Qy      248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlySerGlu 267
Db      788 AGTACCATTTTACCTTCGGGATTAATTATGATGCGGAGATATTAAGAATGTTTCAGAA 847
Qy      268 ThrProLysLeuValSerArgValLeuGlnTrpLysGlnLysProGlnGlnSerSer 287
Db      848 ACAGTAAACTGGTCCAGAAAGGTAAATAATGTTATGATTCGATATGCCAATAAGCTTG 907
Qy      288 ValValTyrAspGlnLeuAsnSerAlaLeuGlnPheMet-----LysGlnLeu 304
Db      908 AAAATATATACGAAGCTCGATCATGCAATTTCTAATTTTATGATGAGCTATCTTAAC 967
Qy      305 ArgGlnMetArgGlnLysTyrHisSerAspProGlnTyrTyrIleLysGlnLeuAspHis 324
Db      968 GATCCCTTACACAGACTCATGACGATTAACAGCATCAGATATTTGAGCTCTTGAGAG 1027
Qy      325 Ser-----ValGlnProLeuThrValAlaIleLysAsn 335
Db      1028 AATGACTGACTCTGCAAAAGATATCTGAAATCAGAAAGTTAGAGATGCAAGTCCACA 1087
Qy      336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGlnValProIleGlnProAspVal 355
Db      1088 ATTAAACGTTCTTTAGAAAAATTAATTAAGATCTGGTCCGATATGAACTCCCGTA 1147
Qy      356 GlnThrGlnLeuLeuAspArgCysGlnGlnIleLeuProGlyCysValGlyValValPro 375
Db      1148 CAATAGCTTATGATGATGATGATGCAACGCTTAAAGAGGTTCTTACTTGCTTAACT 1207
Qy      376 GlyAlaGlyGlyTyrAspAlaIleAlaValLeuVal-----LeuGlnAsn 390
Db      1208 GGTGCTGTGTGTTATGAGCGCATTCAGTATGATTAAGCAAGATGTTGATCTGGGCT 1267
Qy      391 GlnValGlyAsnPheLysGlnLysThrLeuGlnAsnProAspTyrPheHisAsnValTyr 410
Db      1268 CAACCGCTAATGACAAAGA-----TTTCTTAAGGTTCA 1303
Qy      411 TrpValAspLeuGlnGlnThrGlnGlyValLeuGlnGlnLys---ProGlnAspTyr 429
Db      1304 TGGCTGATGTAACTCAGGCTGATCGGGGTGTTAGAAAGAAAGATCCGAAACTTAT 1363
Qy      430 Ile 430
Db      1364 CTT 1366

```

```

RESULT 16
AAB31037
ID AAB31037 standard; DNA; 13917 BP.
XX
AC AAB31037;
XX
DT 31-MAY-2002 (first entry)
XX
DE Plastid transformation vector pHK04 DNA.
XX
KW mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;
KW isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;
KW transgenic plant; yeast; phosphomevalonate kinase; HMGCoA; HMG; AACT;
KW mevalonate kinase; mevalonate diphosphate decarboxylase; HMG-CoA; MD;
KW acetoacetyl thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;
KW MVK; PKM; HMG-CoA reductase; ds.
XX
OS Chimeric - Saccharomyces cerevisiae.
XX
OC Chimeric - Arabidopsis thaliana.
XX
PN W0200210398-A2.
XX
PD 07-FEB-2002.
XX
PF 31-JUL-2001; 2001WO-US24037.
XX
PR 31-JUL-2000; 2000US-221703P.
XX
PA (HANN/) HAHN F M.
XX
PA (KUEH/) KUEHNLE A R.
XX
PI Hahn FM, Kuehnle AR.
XX
DR WPI; 2002-217122/27.
XX
PT Use of specific genes of mevalonate and isoprenoid biosynthetic
PT pathways, for providing a cell with herbicide or antibiotic resistance,
PT and for providing transformed cells having increased isoprenoid
PT production -
XX
PS Claim 77; Page 161-168; 193pp; English.
XX
CC The invention relates to the use of specific genes of the mevalonate
CC and isoprenoid biosynthetic pathways and inactive gene sites
CC (pseudogene). Genes of the invention are used to enhance biosynthesis
CC of isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP),
CC and isoprenoid pathway derived products in the plastids of transgenic
CC plants and microalgae, for producing herbicide or antibiotic resistant
CC transgenic plants and microalgae, for providing transformed cells with
CC increased isoprenoid production compared to non-transformed cells, and
CC for providing a cell with an inserted polynucleotide sequence encoding
CC one or more products of interest. The present sequence is a plastid
CC transformation vector pHK04 containing operon B DNA which encodes
CC the entire mevalonate pathway. This operon contains 5' cerevisiae orf
CC encoding phosphomevalonate kinase (PMK), mevalonate kinase (MVK),
CC mevalonate diphosphate decarboxylase (MD), acetoacetyl thiolase (AACT)
CC and A. thaliana orf encoding 3-hydroxy-3-methylglutaryl-coenzyme A
CC (HMG-CoA) synthase (HMGs) and HMG-CoA reductase (HMGGR).
XX
SQ Sequence 13917 BP; 3882 A; 2923 C; 3063 G; 4049 T; 0 other;
XX
Alignment Scores:
Pred. No.: 2.59e-72 Length: 13917
Score: 809.50 Matches: 190
Percent Similarity: 57.70% Conservative: 76
Best Local Similarity: 41.21% Mismatches: 146
Query Match: 36.30% Indels: 49
DB: 24 Gaps: 14

```

US-10-069-062-7 (1-432) x AAB31037 (1-13917)

3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyGlyTyrLeuValLeuGlnPro 22

```

Db 4141 AAGAGCTTCAGTCCAGGAGAAAGCGTTACTAGCTGCGATATTAGTTAGATTACATA 4200
    |||
Qy 23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProlys 42
    |||
Db 4201 AAATATGAAGCATTTGTAGTCGATTAATGCGAAGATGATGCTGTAGCCCATCTTAC 4260
    |||
Qy 43 GlyThrSerLeuLysLeuSer-----ArgIleLysIleSerSerProGlnPheAla 59
    |||
Db 4261 GGT---TCAATTGCAAGGCTCTGATAAGTTGAAGCGTGTAAAGAACATTAA 4317
    |||
Qy 60 AsnGlyValTyrPgluTyrHisIleSerSerAsnThrGlu---LysProArgGluValGln 78
    |||
Db 4318 GATGGGAGATGGCTGTACATATTAAGTCTTAAAGTGGCTTCACTCTGTTGATAGAGC 4377
    |||
Qy 79 SerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnPro 98
    |||
Db 4378 GGAATCTAAGAACCCCTTCAATGAAAAAGTTATGCTTAACGTAATTTAGCTTAACCT 4437
    |||
Qy 99 Thr-----GluAlaPheAspLeuGluIleIleIleTyrSerAsp 111
    |||
Db 4438 AACATGAGAGCATCTCAATAGAAACCTGTTGTTATGAT-----ATTTCCTGAT 4491
    |||
Qy 112 ProGlyTyrHisSerGlnGluAspThrGluThrLysThrSerSerAsnGlyGluLysThr 131
    |||
Db 4492 GATGCTACCATCTTCAGAGAGATAGCGTTACCGAA-----CATGCTGGCAACAGAGA 4545
    |||
Qy 132 PheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAla 151
    |||
Db 4546 TTGAGTTTCATTCGACAGAAATGAAAGATGCCAAGATCCCAAGAGGCTGGGCTCTCGGCA 4605
    |||
Qy 152 GlyLeuValSerValValAlaThrSerLeuLeuSerHisPheIle-----ProAsn 168
    |||
Db 4606 GGTATATGATACAGTTTAACTACAGCTTGGCCCTCTTTTGTATCGAGCCGTGAATA 4665
    |||
Qy 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
    |||
Db 4666 AATGTAGCAAAATATAGAGAAAGTTATTCATTAATTGACACAACTCTCATGTCAAGCT 4725
    |||
Qy 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
    |||
Db 4726 CAGGCTAAATGGAAGCGGATTGATGAGCGCGCGACATATGATCTATACATAT 4785
    |||
Qy 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGlu 228
    |||
Db 4786 AGAAGATTCACACCCGATTAATCTTAATTCGCCAGATAT-----GGAATGCT 4836
    |||
Qy 229 LysPheProThrGluLeuLysLysLeuIle---GluSerAsnTyrGluLysHisGlu 247
    |||
Db 4837 ACTTACGCGACATTAACCTGGCGCTTGTGATGAAGAAGACTGGAATATTACGATTAA 4896
    |||
Qy 248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlySerGlu 267
    |||
Db 4897 ACTAACCATTTACCTTCGGGATTAACTTTATGATGGCGGATATTAAGATGGTTACAGA 4956
    |||
Qy 268 ThrProLysLeuValSerArgValLeuGlnTyrLysGlyLysGluSerSer 287
    |||
Db 4957 ACAGTAAACTGCTCCAGAAAGGTAAAAATGGTATGATTCGATATGCCAAGAACTTG 5016
    |||
Qy 288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
    |||
Db 5017 AAAATTTATATACAGAACTCATGCAATTTCTAGATTATGATGAGCATATCTAAACTA 5076
    |||
Qy 305 ArgGluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAsn 324
    |||
Db 5077 GATCGCTTACAGAGACTATGATGATTAACAGGACATATTTAGTCTCTGAGAG 5136
    |||
Qy 325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
    |||
Db 5137 AATGACTGATCTGCAAAAGTATCTGAATATCAGAAAGATTAGATGCAAGTGGCACA 5196
    |||
Qy 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspVal 355
    |||

```

```

Db 5197 ATTAGACGTTCTTAGAAAAATTAACATAAGATCTGTCGCGATATCGAACTCCGTA 5256
    |||
Qy 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
    |||
Db 5257 CAATAGCTTATTTGATGATATTCAGACCTTAAAGAGATTTACTTAACTTAACT 5316
    |||
Qy 376 GlyAlaGlyGlyTyrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
    |||
Db 5317 GGTGCTGGGTTATGACCGCATTCAGATTAAGCAAGATGATGATCTAGAGCT 5376
    |||
Qy 391 GlnValGlyAsnPheLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
    |||
Db 5377 CAACCGCTTAAGACAAAAGA-----TTTCTTAAGGTCA 5412
    |||
Qy 411 TrpValAspLeuGluGluGlnThrGluGlyValLeuGluGluLys---ProGluAspTyr 429
    |||
Db 5413 TGGCTGATGATTAAGTACGAGCTGCTGCGGTGTTAGAAAGAAAGATCCGAACTTAT 5472
    |||
Qy 430 Ile 430
    |||
Db 5473 CTT 5475
    |||
RESULT 17
AAD31039
ID AAD31039 standard; DNA; 14623 BP.
AC AAD31039;
XX
DT 31-MAY-2002 (first entry)
XX
DE Placid transformation vector pHK08 DNA.
XX
Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;
KW isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;
KM transgenic plant; Yeast; IPP isomerase; ID11; ds.
XX
OS Chimeric - Saccharomyces cerevisiae.
OS Chimeric - Arabidopsis thaliana.
OS Chimeric - Schizosaccharomyces pombe.
XX
PN WO200210398-A2.
XX
PD 07-FEB-2002.
XX
PF 31-JUL-2001; 2001WO-US24037.
XX
PR 31-JUL-2000; 2000US-221703P.
XX
PA (HANN/) HANN F M.
PA (KUEHN/) KUEHNLE A R.
XX
PI Hahn FM, Kuehnle AR;
XX
DR WPI; 2002-217122/27.
XX
PT Use of specific genes of mevalonate and isoprenoid biosynthetic
PT pathways, for providing a cell with herbicide or antibiotic resistance,
PT and for providing transformed cells having increased isoprenoid
PT production -
XX
PS Claim 77; Page 173-181; 193pp; English.
XX
CC The invention relates to the use of specific genes of the mevalonate
CC and isoprenoid biosynthetic pathways and inactive gene sites
CC (pseudogene). Genes of the invention are used to enhance biosynthesis
CC of isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP)
CC and isoprenoid pathway derived products in the plastids of transgenic
CC plants and microalgae, for producing herbicide or antibiotic resistant
CC transgenic plants and microalgae, for providing transformed cells with
CC increased isoprenoid production compared to non-transformed cells, and
CC for providing a cell with an inserted polynucleotide sequence encoding
CC one or more products of interest. The present sequence is placid
CC transformation vector pHK08 containing operon G DNA which encodes the

```

CC entire mevalonate pathway and ID1 (IPP isomerase). This operon contains
 CC S. cerevisiae, A. thaliana and S. pombe DNA.

CC Sequence 14623 BP; 4128 A; 3028 C; 3227 G; 4240 T; 0 other;

Alignment Scores:

Pred. No.:	2.78e-72	Length:	14623
Score:	809.50	Matches:	190
Percent Similarity:	57.70%	Conservative:	76
Best Local Similarity:	41.21%	Mismatches:	146
Query Match:	36.30%	Indels:	49
DB:	24	Gaps:	14

US-10-069-062-7 (1-432) x AAD31039 (1-14623)

```

QY 3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyLysTyrLeuValLeuGluPro 22
DB 2625 AAGACCTTCAAGTCCCGAGGAGGAGGCTTACTAGCTGGATATTAGTTTGAATACA 2684
QY 23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys 42
DB 2685 AATAATGAAGCATTTGTAGTCGGATTATCGGCAAGATGCAATGCTGTACCCATCTTAC 2744
QY 43 GlyThrSerLeuLysGluSer-----ArgIleLysIleSerSerProGluPheAla 59
DB 2745 GGT---TCATTCGCAAGGCTCTGATAAGTTGAAGTGCCTGTGAAAAGTAAACAATTAA 2801
QY 60 AsnGlyLysIleThrGluIleThrIleSerSerAsnThrGlu---LysProArgGluValGln 78
DB 2802 GATGGGAGTGGCTGTACCATTAATAGCTTAAAGGCTTCAATTCCTTCGATAGGC 2861
QY 79 SerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnPro 98
DB 2863 GGATCTAAGAACCTTTTCAATGAAAAAGTTATCGCTAAGCTATTAGCTTTAAACT 2921
QY 99 Thr-----GluAlaPheAspLeuGluIleIleIleTyrSerAsp 111
DB 2922 AACATGACGACTACTGCAATGAAAACCTTGCTGTAATGAT-----ATTTCCTGAT 2975
QY 112 ProGlyTyrHisSerGlnLysAspThrGluThrLysThrSerSerAsnGlyLysIleThr 131
DB 2976 GATGCCCTTACATTCAGAGAGATAGCGTTACCGA-----CATCGTGCACACAGAGA 3029
QY 132 PheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLysSerAla 151
DB 3030 TTGAGTTTTCATTCGACAGATTTGAAGAAGTTCCCAAAAACGGCTGGCTCTCGGCA 3089
QY 152 GlyLeuValSerValAlaThrSerLeuLeuSerHisPheIle-----ProAsn 168
DB 3090 GGTTTAGTCACAGTTTAACTACAGCTTGCGCTCTTTTGTATCGGACCTGAAAAT 3149
QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisGlySerTyrAla 188
DB 3150 AATGTAGACAATATAGAGAGATTTATCTCAATATTAGCAACAAGTTCCTATTGTCAACT 3209
QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
DB 3210 CAGGCTTAAATTTGAAGCGGTTTATGTAGCGGCGGACGATATGATCTATCAGATAT 3269
QY 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGlu 228
DB 3270 AGAAGATTCACACCGCATTAATCTCAATTATTCACAGATAT-----CGAAGTGCT 3320
QY 229 LysPheProThrGlnLeuLysLysLeuIle---GluSerAsnTyrGluLysHisGlu 247
DB 3321 ACTTACGGCAGTAACTGGCGCATTTGTGTAGTAAAGAACTGGAATATTATTCAGATTAA 3380
QY 248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlyLysSerGlu 267
DB 3381 AGTAAACATTTACCTTCGGGATTAATCTTATGTAGATGCGGATATTAAAGATGTTCAAA 3440
QY 268 ThrProLysLeuValSerArgValLeuGlnIleTyrLysGluLysProGluGluSer 287
  
```

```

DB 3441 ACAGTAAACCTGTCAGAGAGTAAATAATTGTATGATTCGCATATGCCAGAAAGCTTG 3500
QY 288 ValValIleTyrAspGlnLeuAsnSerAlaLeuGlnPheMet-----LysGluLeu 304
DB 3501 AAAATATATACAGAACTGCATGATGCAATTTTATGATTTATGAGATGATTTAACTA 3560
QY 305 ArgGluMetArgGlyLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHis 324
DB 3561 GATCCTTACACAGACATCATGACATTAACAGCATCAGATGATTTGAGTCTCTTGAGAG 3620
QY 325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
DB 3621 AATGACTGTACTCTGCAAAAGATATCTGAAATGACAGACAGATGAGATGCAAGTCCACA 3680
QY 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGlnProAspVal 355
DB 3681 ATTTAAGCTTCTTTTGAATAATTAAGATTTGTTGCGGATATGAACTCCCGTA 3740
QY 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
DB 3741 CAACTAGCTTATTTGATGATGATTCGACAGACCTTAAAGAGATTTCTACTTGTATTA 3800
QY 376 GlyAlaGlyGlyTyrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
DB 3801 GGTGCTGGTGTATGAGCGCATTCGATGATTAAGCAAGATGTTGATCTTACGGCT 3860
QY 391 GlnValGlyAsnPheLysGlyLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
DB 3861 CAACCGCTTAATGACAAAGA-----TTTCTTAAGGTTCAA 3896
QY 411 TrpValAspLeuGluGlnGlnThrGlnGlyValLeuGlnGluLys---ProGluAspTyr 429
DB 3897 TGGCTGATGTATCACTCAGCGCTGACCTGGGCTGTAGGAAGAAAAGATCCGAAACTTAT 3956
QY 430 Ile 430
DB 3957 CTT 3959
  
```

RESULT 18

AAD31041

ID AAD31041 standard; DNA; 14623 BP.

XX AAD31041;

AC AAD31041;

XX AAD31041;

DT 31-MAY-2002 (first entry)

XX 31-MAY-2002 (first entry)

DE Plasmid transformation vector pBK6 DNA.

XX Mevalonate; isopentenyl diphosphate; herbicide resistance; DMAPP; IPP;

XX isoprenoid; dimethylallyl diphosphate; antibiotic resistance; enzyme;

XX transgenic plant; yeast; phosphomevalonate kinase; HMGR; HMGs; AACT;

XX mevalonate kinase; mevalonate diphosphate decarboxylase; HMG-CoA; MDD;

XX acetoacetyl thiolase; 3-hydroxy-3-methylglutaryl-coenzyme A synthase;

XX MVK; PKM; HMG-CoA reductase; IPP isomerase; IPP1; ds.

XX Chimeric - Saccharomyces cerevisiae.

OS Chimeric - Arabidopsis thaliana.

OS Chimeric - Streptomyces sp.

OS Chimeric - Rhodobacter capsulatus.

XX W0200210398-A2.

XX 07-FEB-2002.

PD 31-JUL-2001; 2001WO-US24037.

PF 31-JUL-2000; 2000US-221703P.

PR 31-JUL-2000; 2000US-221703P.

XX (HANN/) HANN F M.

XX (KUEHN/) KUEHNLE A R.

XX Hahn FM, Kuehnle AR;

DR WPI; 2002-217122/27.

XX Use of specific genes of mevalonate and isoprenoid biosynthetic
PT pathways, for providing a cell with herbicide or antibiotic resistance,
PT and for providing transformed cells having increased isoprenoid
PT production

XX Claim 77; Page 185-193; 193pp; English.

XX The invention relates to the use of specific genes of the mevalonate
CC and isoprenoid biosynthetic pathways and inactive gene sites
CC (pseudogene). Genes of the invention are used to enhance biosynthesis
CC of isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP)
CC and isoprenoid pathway derived products in the plastids of transgenic
CC plants and microalgae, for producing herbicide or antibiotic resistant
CC transgenic plants and microalgae, for providing transformed cells with
CC increased isoprenoid production compared to non-transformed cells, and
CC for providing a cell with an inserted polynucleotide sequence encoding
CC one or more products of interest. The present sequence is plasmid
CC transformation vector pKMO6 containing operon B DNA which encodes
CC the entire mevalonate pathway. This operon contains S. cerevisiae orfs
CC encoding phosphomevalonate kinase (PMK), mevalonate kinase (MK),
CC mevalonate diphosphate decarboxylase (MDD), acetoacetyl thiolase (AACT)
CC and A. thaliana orfs encoding 3-hydroxy-3-methylglutaryl-coenzyme A
CC (HMG-CoA) synthase (HMGCS); Streptomyces sp. Cl190 orf encoding HMG-CoA
CC reductase (HMGRT) and R. capsulatus DNA encoding IPP isomerase (IPPI).

XX Sequence 14623 BP; 4128 A; 3028 C; 3227 G; 4240 T; 0 other;

Alignment Scores:

Pred. No.: 2,78e-72 Length: 14623
Score: 809.50 Matches: 190
Percent Similarity: 57.70% Conservative: 76
Best Local Similarity: 41.21% Mismatches: 146
Query Match: 36,30% Indels: 49
DB: 24 Gaps: 14

US-10-069-062-7 (1-432) x AAD31041 (1-14623)

QY 3 LysAlaPheSerAlaProGlyValAlaPheLeuAlaGlyValTyrLeuValLeuGluPro 22
DB 2625 AGAGCCCTCAGTCCCGAGGAAAGCGTTACTAGCTGGAGATATTTAGTTGATAC 2684
QY 23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys 42
DB 2685 AATAATGAGCACTTTGATGCGATTATCGCGAAGATGATCTGTAGCCCATCTTAC 2744
QY 43 GlyThrSerLeuLysGluSer-----ArgIleLysIleSerSerProGlnPheAla 59
DB 2745 GGT---TCATTGCAAGGCTCTGTATAGTTGAAGTCGCTGTAAGAAAGTAAACAATTAA 2801
QY 60 AanglyGluTyrGluLysIleSerSerSerSerSerSerSerSerSerSerSerSerSer 78
DB 2802 GATGGGAGTGGCTGTACCATATTAAGTCTTAAAGGCTCTCTGTTGATGAGGC 2861
QY 79 SerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnPro 98
DB 2862 GCATCTAAGAACCTTTTCATGAAAAAGTTATGCGTAACCTATTAGCTTAAACCT 2921
QY 99 Thr-----GluAlaPheAspLeuGluIleIleIleIleIleIleIleIleIleIle 111
DB 2922 AACATGAGCACTACTGCAATAGAAACTGTTGTTATTCAT-----ATTTCCTCTGAT 2975
QY 112 ProGlyTyrHisSerGlnGluAspThrGluThrLysThrSerSerSerSerSerSerSer 131
DB 2976 GATGCTTACCATTTCTGAGAGATAGCGTTCAGAA-----CATCGTCAACAACAAGA 3029
QY 132 PheLeuTyrHisSerArgAlaIleThrGluValGluIleThrGlyLeuGlySerSerAla 151
DB 3030 TTGAGTTTTCATTCGACAGATTTGAAGAGTTCCAAAAAGGCTGGCTCTCTCGCA 3089
QY 152 GlyLeuValSerValValAlaThrSerLeuLeuSerHisPheIle-----ProAsn 168
DB 3089

DB 3090 GGTATTAGTCACAGTTTAACTACAGCTTTGGCTCTCTTTTGTATCGGACCTGGAAT 3149
QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
DB 3150 AATGTAGACAAATATATAGAGAGATTTATCTAATATATAGACAGATGCTCATTTGCAAGCT 3209
QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
DB 3210 CAGGGTAAATTTGGAAGGGGTTTATGATGAGCGGCGAGCATATGATCTATCATGATAT 3269
QY 209 ArgGlnPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGlu 228
DB 3270 AGAAGATTTCCACCCGCAATTATCTCTAATTTCCAGATAT-----GGAAGTCT 3320
QY 222 LysPheProThrGluLeuLysLysLeuIle-----GluSerAsnTyrGluLysIleGlu 247
DB 3321 ACTTACGGCAGTAACCTGGCCATTTGGTTGATGAGAGAGACTGGAATTTACGATTAA 3380
QY 248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlySerGlu 267
DB 3381 AGTAACATTTTACCTTGGGATTAATCTTTATGATGGGCGATTAAGATGTTCAAGA 3440
QY 266 ThrProLysLeuValSerArgValLeuGlnTyrLysLysGluLysProGluLysSer 287
DB 3441 ACAGTAAACCTGGTCAGAGGTAAATAATGTATGATTCGATATGCGAAGAAAGCTTG 3500
QY 288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
DB 3501 AAATATATATACAGAACTGATCATGCAATTTTATGATATGATGAGCATATCTAACTA 3560
QY 305 ArgGluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHis 324
DB 3561 GATCGCTTACAGAGACTCATGATGATTAAGAGCATGATATTTGAGCTCTTGAAGAG 3620
QY 325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
DB 3621 AATGACTGATCTGTCGAAAGATCTGAAATACAGAAAGTAAAGATGAGATGCTCCACA 3680
QY 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspVal 355
DB 3681 ATTTAGACCTTCTTTAGAAAAATTAAGATCTGTGCGGATTTGAACTCCCGTA 3740
QY 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
DB 3741 CAATAGCTTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3800
QY 376 GlyAlaGlyGlyTyrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
DB 3801 GGTGCTGGTGGTTATGAGCCCATTCAGATGATTAAGCAAGATGATGATGATGATGAT 3860
QY 391 GlnValGlyAsnPheLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
DB 3861 CAACCCGCTAATGACAAAGA-----TTTCTAAGTTTCA 3896
QY 411 TrpValAspLeuGluGlnGlnThrGluGlyValLeuGluGluLys-----ProGluAspTyr 429
DB 3897 TGGCTGATGATTAATCAGAGCTGATCGGGGTGTTAGAAAGAAAGATCCGAAACTAT 3956
QY 430 Ile 430
DB 3957 CTT 3959
RESULT 19
ABT19236
ID ABT19236 standard; DNA; 1356 BP.
XX ABT19236;
AC ABT19236;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene #1594.
XX
XX Fungicide; cyrostatic; essential gene; Aspergillus fumigatus; infection;
KW


```
DB 730 GCTGCATTACGGCTCTGCTCTCTACAGACGCTTCCCTCCATTCGAATCAGTG 789
QY 220 -----PheGlnValLeuGlnSerAspPro 227
DB 790 GGTGACCGCTGTTCTCCAGGCTTCGAAGACGGTGTCTGAGAGACGCCGAC 849
QY 228 GlnTyrPheProThrGlnLeuLysLeuLeuLeuSerAsnTrpGlnGluLysHisGln 247
DB 850 CCCACGATCCG-----TGGACACCCGAA----- 873
QY 248 ArgCys-----ThrLeuProTyrGlyLysLeuLeuMetGlyAspVal 262
DB 874 ---TGTCTAGACTTCGGCATGAGACTCCCGCGGAAATCAATGCTCTGTGACGTC 930
QY 263 LysGlyGlySerGlnThrProLysLeuValSerArgValLeuGlnTrpLysGlyLys 282
DB 931 GAATGTGTTCGAGATCCATCCATGAGAGAAAGTTTGGAGTGGCGAAACAGAA 990
QY 283 ProGlnLysSerValValTyrAspGlnLeuSerAlaAsnLeuGlnPheMetLys 302
DB 991 CAGAAGAAAGCCGATATGCTCTCTGGGGCGCTTCGATCCAGACAGAGACTTCCCTG 1050
QY 303 GlnLeuArgGlnMetArgGlnLysTyrAspSerAspProGlnThrTyrLysGlnLeu 322
DB 1051 GAATCAGACGCTTGGCAGACGCCGAC----- 1080
QY 323 AspHisSerVal-----GluProLeuThrValAlaLeuLysAsnLeuArgLysGly 339
DB 1081 GAACATACCTCTCAGTCACTTGAATAATGCCGACCTATATTCAGCGCTCGCTAACAC 1140
QY 340 LeuGlnAlaLeuThrGlnLysSerGlnValProLeuGlnProAspValGlnThrGlnLeu 359
DB 1141 ATCCCTTCATGATCAAAAGTCGATGCCATGATGACCCGGCTCCAAACCGAGCTA 1200
QY 360 LeuAspArgCysGlnGlnLeuProGlyCysValGlyValValProGlyLysGly 379
DB 1201 CTGTGACGCTCTGTCGAGCTGAGGGGTGCATCGGGTGGTGGTTCAGAGAGAGGGGCG 1260
QY 380 TyrAspAlaIleAlaValLeuValLeuGlnAsnGln-----ValGlyAsnPheLysGln 397
DB 1261 TACGAGCCGATGTGCTCTCATCCAGACAAATCCGATGTGATCCAGATTGAAAGCC 1320
QY 398 -----LysThrLeuGlnLysAsn-----ProAspTyr-----PheHisAsnVal 409
DB 1321 TTCCTTGAACCTTGGAGAGCAAAAGCGAGAGAGATTTGGTGGCAAGATTGGAAAGTC 1380
QY 410 TyrTrpValAspLeuGlnGlnGlnThrGlnGlyValLeuGlnGlnLysAspTyr 429
DB 1381 AGGCTTCTGGCGGTCCGATGATCAGAGGAGTCAAGACGAGATGCTCGAGCAATAT 1440
QY 430 IleGly 431
DB 1441 GCGGCG 1446
RESULT 21
ABT18642
ID ABT18642 standard; DNA; 1509 BP.
XX
XX ABT18642;
XX AC
XX AC 16-APR-2003 (first entry)
XX DE Aspergillus fumigatus essential gene #1000.
XX KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
XX KW cancer; contamination; biofilm; antibody; immune response; ds.
XX OS Aspergillus fumigatus.
XX PN WO200286090-A2.
XX PD 31-OCT-2002.
```

```
XX 23-APR-2002; 2002WO-US13142.
PF 23-APR-2001; 2001US-285697P.
PR 27-APR-2001; 2001US-287066P.
PR 05-JUN-2001; 2001US-295890P.
PR 09-JUL-2001; 2001US-303899P.
PR 31-AUG-2001; 2001US-316362P.
XX
XX (ELIT- ) ELITRA PHARM INC.
XX
XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX WPI; 2003-093124/08.
XX
XX New purified or isolated nucleic acids of essential genes of
XX Aspergillus fumigatus, useful for treating or preventing infections by
XX A. fumigatus, or for treating a non-infectious disease in a subject
XX e.g. cancer
XX
XX Disclosure; Page -; 175pp; English.
XX
XX The invention relates to novel purified or isolated nucleic acids of
XX essential genes of Aspergillus fumigatus. The isolated nucleic acids of
XX the invention are used to treat or prevent infections by a pathogenic
XX organism such as A. fumigatus, to treat a non-infectious disease in a
XX subject (e.g. cancer), to prevent or contain contamination of an object
XX by A. fumigatus, or to prevent or inhibit formation on a surface of a
XX biofilm comprising A. fumigatus. The polynucleotides are useful for
XX expressing recombinant protein for characterisation, screening or
XX therapeutic use, as markers for host tissues in which the pathogenic
XX organisms invade or reside, for comparing with the DNA sequence of A.
XX fumigatus to identify duplicated genes or paralogues having the same or
XX similar biochemical activity and/or function, for comparing with DNA
XX sequences of other related or distant pathogenic organisms to identify
XX potential orthologous essential or virulence genes, for selecting and
XX making oligomers for attachment to a nucleic acid array for examination
XX of expression patterns, for raising anti-protein antibodies, as an
XX antigen to raise anti-DNA antibodies or to elicit another immune
XX response, and for identifying polynucleotides encoding the other protein
XX with which binding occurs or to identify inhibitors of the binding
XX interaction. The polypeptides may be used to raise antibodies or to
XX elicit immune response, as a reagent in assays designed to quantitatively
XX determine levels of the protein in biological fluids, as a marker for
XX host tissues in which pathogenic organism invade or reside, and to
XX isolate correlative receptors or ligands in the case of virulence
XX factors. This polynucleotide sequence represents one of the essential
XX genes of Aspergillus fumigatus of the invention.
XX
XX Sequence 1509 BP; 338 A; 418 C; 429 G; 324 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 3.9e-57 Length: 1509
XX Score: 649.50 Matches: 166
XX Percent Similarity: 47.88% Conservative: 83
XX Best Local Similarity: 31.92% Mismatches: 150
XX Query Match: 29.13% Indels: 121
XX DB: 25 Gaps: 16
XX
XX US-10-069-062-7 (1-432) x ABT18642 (1-1509)
QY 4 AlapheserAlaProGlyLysAlaPheLeuAlaGlyLysTyrLeuValLeuGlnProIle 23
DB 28 GCTTGTCCGCGCGGAGAGGTCTCTCACTGAGGGGTATCTGTCTGAGACCGCAGC 87
QY 24 TyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIle----- 39
DB 88 TACACTGGAGCTGTGTTGCCCTTACGACCGCAAGATCCATGTCATGTTCCAGCAATTGGA 147
QY 40 -----ThrProLys 42
DB 148 CGAACAATCGCGGAGAGCGCGCTCGGATTCGCGCATGCGCGGTCCGAGACACACTCGAG 207
```


PD 31-OCT-2002.
XX 23-APR-2002; 2002WO-US13142.
XX
PR 23-APR-2001; 2001US-285697P.
PR 27-APR-2001; 2001US-287066P.
PR 05-JUN-2001; 2001US-295890P.
PR 09-JUL-2001; 2001US-303899P.
PR 31-AUG-2001; 2001US-316362P.
XX
PA (ELIT-1) ELITRA PHARM INC.
PI Jiang B, Tishkoff D, Zamudio C, Eroshtkin AM, Hu W, Lemieux SM;
XX MPI; 2003-093124/08.
DR
XX
PT New purified or isolated nucleic acids of essential genes of
PT Aspergillus fumigatus, useful for treating or preventing infections by
PT A. fumigatus, or for treating a non-infectious disease in a subject
PT e.g. cancer.
XX
PS Disclosure; Page -; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or inhibit formation of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterization, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organisms invade or reside, and to
CC isolate correlative receptors or ligands in the case of virulence
CC factors. This polynucleotide sequence represents one of the essential
CC genes of Aspergillus fumigatus of the invention.
XX
XX Sequence 3508 BP; 857 A; 916 C; 883 G; 852 T; 0 other;
XX
Alignment Scores:
Pred. No.: 1,29e-56 Length: 3508
Score: 649.50 Matches: 166
Percent Similarity: 47.88% Conservative: 166
Best Local Similarity: 31.92% Mismatches: 150
Query Match: 29.13% Indels: 121
DB: 25 Gaps: 16
US-10-069-062-7 (1-432) x ABR18048 (1-3508)
QY 4 AAlpHeSerAlaProGlyLysAlaPheLeuAlaGlyTyrLeuValIleuGluProIle 23
DB 1027 GCTTGTTCGGCGCGGCAAGGCTCTCTCACTGGGGGTTATGTGGCTTGGACCGCAGC 1086
QY 24 TYZAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIle----- 39
DB 1087 TACACGGGAGCTGTGTTGGCTTGAACGACAGATCATGTCTTTCAGCAATTGAGA 1146
QY 40 -----ThrProLys 42
DB 1147 CGGAACCATCGCGGGAAGCGGCTCGGATCCGCGCATGCGCGTCCGACACACTCAG 1206

QY 43 -----GlyThrSerLeuYsgIuSerArgIleYsgIleSerSerPro 56
DB 1207 GCGGAAGCAANTTCATGAGACAGAGAGACAGACAGATGTGTGACACTCCCA 1266
QY 57 GlnPheAlaenGlyGluTyrGluTyrHisIle----- 67
DB 1267 CAGTTCGTGATCGCGTATGAGATATGACATACAAAGATCGAGACGGTGGAGAGTC 1326
QY 67 ----- 67
DB 1327 CTATGAAAACAGAAATGATGGTAAGCAAAATATATATGATCATGAAATTTGCCCA 1386
QY 68 SerSerAntThrGluLeuProArgGluValGlnSerArgIleAsnProPheLeuGluAla 87
DB 1387 GGAGCTGACCCATCGAGCGCAGC-----AACCTGTGTTGCGAGACC 1428
QY 88 ThrIlePheIleValIleuAlaTyrIle-----GlnProThrGluAlaPheAsp 103
DB 1429 TCTGTGAACCTTCGCTTGAATTAATCATGATGATGCGCGCACTCGAAGATTTCCGGTCA 1488
QY 104 LeuGluIleIleIleTyrSerAspProGlyTyrHisSerGlnIuAspThrGluThrLys 123
DB 1489 TTATCGATTACTATCTCGCGCAGACAGATTAATCTCGAGACGGCTTCCAGGCT 1548
QY 124 ThrSerSerAsnGlyGluTyrThrPheLeuTyrHisSerArgAlaIleThrGluValGlu 143
DB 1549 TCGGACATCGGCTCGTCAAGACAGATTCGTGACCTTGTGTTGGCTTCAGAGGACAC 1608
QY 144 LysThrGlyLeuGlySerSerAlaGlyLeuValSerValAlaIleThrSerLeuLeuSer 163
DB 1609 AAGACAGCGCTGGGCTTTCAGCCCGCTTGGCAGCTGCGCTGCTCTCTCTCTCATC 1668
QY 164 His-----PheIleProAsnValIleSerThrAsnLysAspIleLeuHisAsnValAla 181
DB 1669 CACCGTACTATGCAACCGGACATCTCGGTCCAGGCGCGAGACCTTCACATGTGGC 1728
QY 182 GlnIleAlaHisCysTyrAlaGlnLysIleGlySerGlyPheAspValAlaThrAla 201
DB 1729 CAGGGGCGCCATCGCGCTCGCGAGGTAAGTCGGGTGCGGTTCGATGTTGCACCTGCC 1788
QY 202 IleTyrGlyLeuIleValTyrArgArgPheGlnProAlaLeuIleAsnAspVal----- 219
DB 1789 ATTATCGGCTCTGCTCTACAGACGCTTCCCTCGATTCGAATTCAGTGGGTAC 1848
QY 220 -----PheGlnValIleuGlnSerAspProGluLys 229
DB 1849 GCTGTTCTCCAGGCTTGAAGAGCGGTTGTCGGAATCGTAGAGAGCCGACCCACG 1908
QY 230 PheProThrGluLeuLysLeuIleGluSerAsnTyrGluGluLysHisIleGluArgCys 249
DB 1909 CATCG-----TGGACACCGAA-----TGT 1929
QY 250 -----ThrLeuProTyrGlyIleLysLeuLeuMetCysIAspValIleGly 264
DB 1930 CTAGACTTCGCGATGAACCTCCCGCGGAAATGCAAAATGCTCTGTGACGTCGAAATGT 1989
QY 265 GlySerGluThrProLysLeuValSerArgValIleuGlnTyrPylsGluLysProGlu 284
DB 1990 GGTTCGACAGCTTCATCCATGAGTGAAGAGTTTGGAGTGCGGAAACAGAAATCAGAG 2049
QY 285 GluSerSerValTyrAspGlnIleuAsnSerAlaAsnLeuGlnPheMetCysGluLeu 304
DB 2050 GAAGCGGATATGCTTGGGGCGCTTCGAAATCGAACAACAGAGACTTCGCGTGAATC 2109
QY 305 ArgGluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHis 324
DB 2110 AAGACCTTGAGCAGAGCCGAGAC-----GAAAT 2139
QY 325 SerVal-----GlnProLeuThrValAlaIleLysAsnIleArgLysGlyLeuGln 341
DB 2140 ACTCTCACTGACTTTGAAATATGTCGACCTATATTCAGCGCTCGGTAAACACATCCGT 2199


```

QY 182 GlnIleAlaHisCysTyrAlaGlnIleValSerGlyPheAspValAlaThrAla 201
Db 1730 CAGCGCGCCCACTGCGCGCGAGGAAAGTCGGGCTTCATGTTGACGCTCC 1789
QY 202 IleTyrGlyLeuIleValTyrArgArgPheGlnProAlaLeuIleAsnAspVal 219
Db 1790 ATTACGGGCTCTGCTCTACAGACGCTTCCCGCTCCGATTCGATGAGGCTGAC 1849
QY 220 -----PheGlnValLeuIleAsnAspProGluIleVal 229
Db 1850 GCTGCTTCCAGGCTTCGAGAGCGGTTGTCGAGATCGTACAGAGACGCCGAC 1909
QY 230 PheProThrGluLeuIleValSerGlyLeuIleGluSerAsnTrpGluIleValSerGly 249
Db 1910 CATCCG-----TGAGACACCCGA-----TGT 1930
QY 250 -----ThrLeuProTyrGlyIleValSerLeuIleValSerGly 264
Db 1931 CTAGACTTCGCGCATGAAAGCTCCCGCGGAAATGATGCTCTGTGACGTCGATG 1990
QY 265 GlySerGlnIleProIleValSerArgValLeuGlnIleProIleValSerGly 284
Db 1991 GGTTCGACAGCTCATCCATGAGGAAAGTTTGAGTGGCGGAAACGAAATCAGAG 2050
QY 285 GluSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetIleValSerGly 304
Db 2051 GAAGCCGATATGCTCTGCGGCGCTCTGCGATCGAACACAGACGCTCCGCGAACT 2110
QY 305 ArgGluMetArgIleValTyrAspSerAspProGluIleValSerGly 324
Db 2111 AGACGCTTGCGACAGACCCGAC-----GAAAT 2140
QY 325 SerVal-----GluProIleValAlaIleValSerLeuIleValSerGly 341
Db 2141 ACTCTAGAGCTTGAATAATGTCGACCTATATTCAGGCTGCGTAAACCATCCG 2200
QY 342 AlaLeuThrGlnIleSerGlnValProIleGluProAspValGlnIleValSerGly 361
Db 2201 TCCATGACTCAAAAGTCGATGCCATCGAACCGCGCTCCAAACCGACTACTGAC 2260
QY 362 ArgGlyGlnIleProGlyCysValGlyValValProGlyAlaGlyIleValSerGly 381
Db 2261 GCTCTGTCGAGCTGAGGAGTGTGATCGGTGTGTTCCAGAGACGAGGCGCTACG 2320
QY 382 AlaIleAlaValLeuValLeuGlnValGln-----ValGlyAsnPheIleValSerGly 397
Db 2321 GCCATTGTGCTCTCATCCAAAGCAATCCGATGTGATCCAGATGAAAGCTTCTT 2380
QY 398 LysThrLeuGlnIleAsn-----ProAspTyr-----PheHisAsnValTyrTrp 411
Db 2381 GAACTTGGAGAGCAAGAGGAGAGCATTTCCGTGCGCAAGATTGGGAAAGTCAGGCT 2440
QY 412 ValAspLeuGlnIleGlnIleValValLeuGlnIleValSerProGluIleValSerGly 431
Db 2441 CTGCGGCTCGCATGATGATCAGAGGAGTCAAGAGATGCTCGAGCAATATGCGGG 2500

```

RESULT 25

```

ABL60244
ID ABL60244 standard; cDNA; 2396 BP.
XX
AC ABL60244;
XX
DT 29-JUL-2002 (first entry)
XX
DE Arabidopsis thaliana PMVK encoding cDNA SEQ ID NO 1.
XX
KM Thale cress; PMVK; phosphomevalonate kinase; plant; herbicide;
XX
KW growth regulator; enzyme; gene; ss.
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers

```

```

FT CDS 685..2202
FT /tag= a
FT /product= "PMVK"
XX
XX DE10057755-A1.
XX
XX 23-MAY-2002.
XX
XX 22-NOV-2000; 2000DE-1057755.
XX
XX 22-NOV-2000; 2000DE-1057755.
XX
XX (FARB ) BAYER AG.
XX
XX Meisener R, Lechelt-Kunze C;
XX
XX WPI, 2002-445360/48.
XX
XX P-PSDB; ABB77505.
XX
XX New nucleic acid encoding plant phosphomevalonate kinase, useful for
XX identifying modulators, potentially useful as herbicides and growth
XX regulators -
XX
XX Claim 6; Page 10-12; 18pp; German.
XX
XX The invention relates to a nucleic acid (I, ABL60244) that encodes a
XX plant phosphomevalonate kinase (PMVK, ABB77505), excluding the known
XX fully defined partial sequences ABL60245-ABL60247. Plant PMVK, (I),
XX constructs and host cells that contain (I) are used to identify agents
XX that bind to and/or modulate activity of PMVK, potentially useful as
XX herbicides and growth regulators. (I) is also used for recombinant
XX production of PMVK.
XX
SQ Sequence 2396 BP; 669 A; 492 C; 558 G; 677 T; 0 other;

```

Alignment Scores:

```

Pred. No.: 8,36e-39 Length: 2396
Score: 475.00 Matches: 143
Percent Similarity: 43.76% Conservative: 78
Best Local Similarity: 28.32% Mismatches: 184
Query Match: 21.30% Indels: 100
DB: Gaps: 15

```

US-10-069-062-7 (1-432) x ABL60244 (1-2396)

```

QY 1 MetSerLysAlaPheSerAlaProGlyIleValAlaPheLeuAlaGlyTyrIleValLeu 20
Db 685 ATGGCTGTTGTTGCTTCTGCTCTGCGAAAGTTTGATGACTGAGGCTACCTTGACTC 744
QY 21 GluProIleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThr 40
Db 745 GAGAGCCAAATGACGAGGCTTGTGTGATCAAAATGACGCTTTTACCGCATTTGTAAG 804
QY 41 ProLysGlyThrSerLeuIleGlu-----SerArgIleLysIle 53
Db 805 CCAATCAAGAGAGAGATCAAGCTGGAAGTGGCAGTGAAGATGACAGATGCAAAATTA 864
QY 54 SerSerProGlnPheAlaAsnGlyIleTrpGlyIleValSerSerAsn-----Thr 71
Db 865 ACATCAACCAAGCTCTCGAGA---GAAAGCATGTATTAACCTGTCACTGATCATTTGACT 921
QY 72 GluLysProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThrIlePheIle 91
Db 922 CTTCAGTCTGTGTCGCAAGTATTCAGAAACCCCTTTGAGACATGCGATACAGTAT 981
QY 92 ValLeuAlaTyrIleGlnProThrGluAlaPheAsp----- 103
Db 982 GCTATAGCTGCTGCTCATTTGGCAACGAGAGCAAGAAATCATTCGCAAACTCTTA 1041
QY 104 -----LeuGlnIleIleIleTyrSerAspProGlyIleValSerGlnIleValSerGly 119
Db 1042 TTGCAAGCTTGTATTAACAATATTTAGGCTCCAAATGACTTTTACTCATATGCAACG 1101

```


CC IPP pathway is produced. This polynucleotide sequence represents a
 CC nucleic acid sequence encoding an acetyl-coA acetyltransferase enzyme, a
 CC variant of one of the enzymes used to synthesize IPP from acetate.

XX Sequence 1509 BP; 441 A; 299 C; 357 G; 412 T; 0 other;

Alignment Scores:

Pred. No.:	4,87e-36	Length:	1509
Score:	445.50	Matches:	138
Percent Similarity:	44.84%	Conservative:	88
Best Local Similarity:	27.38%	Mismatches:	179
Query Match:	19.98%	Indels:	99
	25	Gaps:	16

US-10-069-062-7 (1-432) x AAL55372 (1-1509)

```

QY 1 MetSerLysAlaHisSerAlaProGlyLysAlaPheLeuAlaGlyLysLysLeuValLeu 20
DB 1 ATGGCTGTAGTCTCTGCTCCGGGTAAAGTGTGATGACTGGGGGTAACTCATATG 60
QY 21 GluProIleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThr 40
DB 61 GAAGACCCCATGACAGGATTTGATCTCAGCAAAATGCTGATTTCTATGCCATTGTGAAG 120
QY 41 ProLysGlyThrSerLeuLysGlu-----SerArgIleLysIle 53
DB 121 CTTATTTAGATGAATCAAACTGATAGTTGGGCATGGGCATGCATGCATGTGAAATTA 180
QY 54 SerSerProGlnPheAlaAsnGlyLysLysLysLysLysLysLysLysLysLysLys 73
DB 181 ACATCTCCCACTACATACAGG---GAAGCTGTGACAAATGCTGACGTGAAATTTAGCT 237
QY 74 ProArgLysValLeuGlnSerArgIle-----AsnProPheLeuGlnAlaThrIlePheIle 91
DB 238 CTTGACGTGTCTCTTCAATGATCATGATCAAGAACCCATTGTGAAACAGACATGCAATTT 297
QY 92 ValLeuAlaTyrIleGlnProThrGlnAlaPheAsp----- 103
DB 298 GCTGTAGACGTGACATGACATCACTTGAACAAAGATAGAAAGATGCTTAAACAGCTA 357
QY 104 -----LeuGlnIleIleIleTyrSerAspProGlyLysLysLysLysLysLysLys 119
DB 358 CTCTTGCAAGCTGTGATATTAACAATATAGTACCAATGACTTCAATCATACCGAAT 417
QY 119 ----- 119
DB 418 GAGATTGAAGCATGTGACTCCCTTGCACCCGAATCATGTGCTGCATCTCTTCTTT 477
QY 120 -----ThrGluThrLysThrSerSerAsnGlyLysLysLysLysLysLysLys 136
DB 478 TCCTCATATCACCTTCAATGATAGAGAAAGAAATGACAA----- 516
QY 137 ArgAlaIleThrGlnValGlyLysThrGlyLeuGlySerSerAlaGlyLeuValSerVal 156
DB 517 AACTGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 576
QY 157 ValAlaThrSerLeuSerLeuSerHisPhe-----IleProAsnValIleSerThrAsnLys 174
DB 577 GTAGTGTGCTTCTTACTTCACTCCTTGGATGTGATGATGATGATGATGATGATGATGAT 636
QY 175 -----AspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
DB 637 AAGAAATTTTCTGATCTGTATTTGATGATGATGATGATGATGATGATGATGATGATGAT 696
QY 189 GlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 208
DB 697 CAAGGAAAGTCCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 756
QY 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGlnLysSerAspProGln 228
DB 757 GTGCGCTTCTCTCAGAGAGCTTCTCTCTCAGAGAGCTGGAAGAGATTTCA--- 813
QY 229 LysPheProThrGlnLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLys 248
  
```

```

DB 814 ---TTACAGAGAGTATTCTTAACATCTTAAGAAATGGAGACATAGAGACTATG 870
QY 249 CysThrLeuProTyrGlyLysLeuMetGlyAspVal---LysGlyLysGlu 267
DB 871 TTTCTCTGACCACTGATGAGCTGTACTAGGTGAGCAGAGAACTGAGAACTCTTCC 930
QY 268 ThrProLysLeuValSerArgValLeuGlnIleTyrLysGluLysProGlnLysSer 287
DB 931 AGCCATCATATGTTAGTCTCTTAAGAAATGCAAGATGATGATGATGATGATGATGATGAT 990
QY 288 ValValTyrAspGlnLeuLysSerAlaAsn-----LeuGlnPheMet 301
DB 991 GAACATGAGAAAGTGTCAAGCGCAAAATTCAGACCTTGAACCCAAATTCATATTTTA 1050
QY 302 LysGluLeuArgGluMetArgLysLysTyrAspSer----- 313
DB 1051 ACCAAGCTCGCAAG---GAAACATGGGAGCGGTAAATGTTGATAGACAGTTGC 1104
QY 314 AspProGluThrTyrIleLysGluLeuAspHisSerValGluProLeuThrValAlaIle 333
DB 1105 ACCACAAAATACTCAGAGAGATGATGACGACGACATGAAACCCAGCAGAACAGT 1164
QY 334 -----LysAsnIleArgLysGlyLeuGlnAla 342
DB 1165 GTTAAAGCATTTATGATCAAGAAATGCCATGCTTCAGATCAAGAAATTAATGCGCCAG 1224
QY 343 LeuThrGlnLysSerGlnValProIleGluProAspValGlnThrGlnLeuLysAspArg 362
DB 1225 ATGGGTAGAGCTGACAGGTGTTCCGATAGAGCTGTGAATCAACATCTGATTTGATCT 1284
QY 363 CysGlnGluIleProGlyCysValGlyLysValProGlyAlaGlyLysTyrAspAla 382
DB 1285 ACTATGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1344
QY 383 IleAlaValLeuValLeuGlnLysGlnValGlyAsnPheLysGln-----LysThrLeu 400
DB 1345 GTCTTGTGCTTACCTTGGGAGCTGTGTACCAATGTGGCAAAAGCTTGGAGTTCACTC 1404
QY 401 GluAsnProAspTyrPheHisAsnValTyrTrpValAspLeuGlnLysLysLysLys 420
DB 1405 -----AsnGlyTGTGCTGTGTTAGAGAAAGCCCTAATGCT 1443
QY 421 ValLeuGlnLys 424
DB 1444 GTTTGTTAGAA 1455

RESULT 27
AB251452
ID AB251452 standard; cDNA; 869 BP.
XX
AC AB251452;
XX
DT 28-MAR-2003 (first entry)
XX
DE Aspergillus oryzae polynucleotide SEQ ID NO 565.
XX
KW Aspergillus oryzae; fermentation; fungus; industrial; EST;
XX expressed sequence tag; gene; ss.
XX
OS Aspergillus oryzae.
XX
PN MO200279476-A1.
XX
XX 10-OCT-2002.
XX
XX PD 22-MAR-2002; 2002WO-1B00890.
XX
XX PF 30-MAR-2001; 2001JP-0098371.
XX
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX (NARB-) NAT RES INST BREWING.
XX (NORO) NAT FOOD RES INST MIN AGRIC.
  
```


XX Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
 PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
 XX MPI, 2003-046817/04.
 XX
 PT Detection of expression of specific *Aspergillus* genes for monitoring
 PT the fermentation and growth conditions of the fungus, using DNA probes
 XX
 PS Claim 1; SEQ ID NO 565; 48bp + Sequence listing; Japanese.
 XX
 CC The invention relates to a polynucleotide having any of 6006 specific
 CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under
 CC specific culture conditions including one or more of eutrophic,
 CC oligotrophic, solid, early germination, alkaline, high temperature, low
 CC temperature or maltose culture or polynucleotides stringently hybridizing
 CC to these sequences. The polynucleotides are useful for monitoring the
 CC progress of fermentation and the growth conditions of a fungus,
 CC especially of *Aspergillus oryzae* which is widely used in industrial
 CC fermentation. Also monitoring for fungal contamination.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 869 BP; 193 A; 226 C; 239 G; 211 T; 0 other;

Alignment Scores:

pred. No.: 1,75e-30 Length: 869
 Score: 388.50 Matches: 97
 Percent Similarity: 52.50% Conservative: 50
 Best Local Similarity: 34.64% Mismatches: 89
 Query Match: 17.42% Indels: 44
 DB: Gaps: 9

US-10-069-062-7 (1-432) x ABZ51452 (1-869)

QY 47 LysGluSerArgIleIleSerSerProGlnPheAlaAsnGlyGluTrpGluTyrHis 66
 DB 67 AAGAGAGAGCGTGGTGTGTACCGCTCCGCAATTTGTCAATGGCAATTTGGAGTACGCT 126
 QY 67 Ileser-----SerAnthrGluLysProArgGluValGln-----SerArgIle 81
 DB 127 ATACACGCTGTGAGATGAGAGGTGAGATCAATCAATCAAGAGAGAGCGGCTGCC 186
 QY 82 AAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIle-----Gln 97
 DB 187 AATCGCTGTGCAAACTTCCTCACTACGCTTCACTATATATGCTATGTGCCGAC 246
 QY 98 ProThrGluAlaPheAspLeuGluIleIleIleTyrSerAspProGlyTyrHisSerGln 117
 DB 247 TCGAAGAGACTTGGGTCCTCTCCGACCATCTGCGGACACATGATTAATCTCTGAG 306
 QY 118 GluAspThrGluThrIleTyrSerSerAnGlyGluLysThrPheLeuTyrHisSerArg 137
 DB 307 ACTGCTCTCTAGGGTTCTGTAGTCCCTGGAAGA-----TTGCTGAACCTTCGTTT 360
 QY 138 AlaIleThrGluValGluLysThrGlyLeuGlySerSerAlaGlyLeuValSerValVal 157
 DB 361 CCTCTTCAGAGGCCCAAGACAGACATGATCTCTGCGCTCTACTACTACCTGCCCTTA 420
 QY 158 AlaThrSerLeuLeuSerHis-----PheIleProAsnValIleSerThrAsnLysAsp 175
 DB 421 GATATCTCTCTGTTTTCACCGTACCTCCGACCTGACAGACACTTGGACCTTCTGTGAC 480
 QY 176 IleLeuHisAsnValAlaGlnIleAlaHisCysTyrAlaGlnLysIleGlySerGly 195
 DB 481 AAGCTTCACTATCTGGCAGAGCGCTGCTGCTCAAGATTAAGTGGATCCGGG 540
 QY 196 PheAspValAlaThrAlaIleTyrGlyLeuIleValTyrArgArgPheGlnProAlaLeu 215
 DB 541 TTGTATGTGCTGCTGCTATCTAAGGCTCTTGGCTATATATGCGGATTCCTCCCAAGCAT 600

QY 216 IleAsnAspVal-----PheGlnValLeu 223
 DB 601 TTGGAATCCGTGGGAGAGCGGGTTCACTGGGTTTGAAGAGCGGCTGTTGACATGCTG 660
 QY 224 GluSerAspProGluLysPheProThrGluLeuLysLeuIleGluSerAnTrpGlu 243
 DB 661 GAGGACCGCCGACCTTACGATCCG-----TGGGAT 690
 QY 244 GluLysHisGluArgCys-----ThrLeuProTyrGlyIleLysLeu 258
 DB 691 ACAGAG-----TGCCTGATTTCCGATCGACATCTTCGAGGCAATGAGTGC 741
 QY 259 MetGlyAspValLysGlySerGlu-ThrProLysLeuValSerArgValLeuGlnTr 278
 DB 742 CTGTGCAATGTTGATGGGGCTCAAAATTCCTTCGATGTCAGAGAAAGTCTGAAATG 801
 QY 278 PLYLysGluLysProGluGluSerSerValTyrAspGlnLeuAsnSerAlaAsn 297
 DB 802 GCGAAACAAACACAGACAGAACCGATCTTCTTGGGCTGCCCTCCAGTCAAAACAT 859
 RESULT 28
 AA08722
 ID AA08722 standard; cDNA; 586 BP.
 AC AA08722;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Fusarium venenatum EST SEQ ID NO:1245.
 XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; *Aspergillus niger*;
 KW *Aspergillus oryzae*; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Fusarium venenatum.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 DR MPI, 2000-594572/56.
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 86; Page 854; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore

CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF1247 represents ESTs from
CC Fusarium venenatum; AAF1248 to AAF1853 represents ESTs from Aspergillus
CC niger; AAF1854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.

XX Sequence 586 BP; 141 A; 143 C; 175 G; 119 T; 8 other;

Alignment Scores:

Pred. No.:	1,41e-18	Length:	586
Score:	271.00	Matches:	67
Percent Similarity:	51.06%	Conservative:	29
Best Local Similarity:	35.64%	Mismatches:	68
Query Match:	12.15%	Indels:	24
DB:	21	Gaps:	4

US-10-069-062-7 (1-432) x AAF08722 (1-586)

```
QY 208 TTTATGATGPhgInProAlaLeuIleAsnAspValPheGlnValLeuGlnSerAspPro 227
DB 7 TACAGAGGCTTCTCTCCGCGACGCTGAC-----AAGATCCCGCGCTGGGCG 57
QY 228 GlnUlyPheProthrGlnUleuUlyLeuIle-----GlnSerAspTrp-----Gln 243
DB 58 GCTGCTCCCTGATGCTTGTGGAAGCTTGTGAGCGAGTCTGAGTGTGAGTGTGAG 117
QY 244 GlnUlyHisGlnUlyArgCysThrLeuProTyrGlyIleUlyLeuUmetGlyAspValUly 263
DB 118 GTTCTCAAGACCGCTGACCATGACCCCAAGGCTGCTTGTGCGTATGTCGATGTTGAC 177
QY 264 GlyGlySerGlnUlyThrProUlyLeuValSerArgValLeuGlnTrpUlyUlyGlnUly 283
DB 178 TGGGGAAGCAAAACCGTCGATGCTCAANAAAGTCTCCGTCGAAAGGCTCAAAAGCC 237
QY 284 GlnGlnSerSerValUlyTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetUlyGln 303
DB 238 GACGAATCCAAAGCCCTTGGATGAACTGCAGAAACGTAACAGCAATTGATGGCACT 297
QY 304 LeuArgGlnUmetArgGlnUlyTyrAspSerAspProGlnUlyThrUlyLeuUlyAsp 323
DB 298 CTTAAGCC----- 306
QY 324 HisSerValGlnProLeuThrValAlaIleUlyAsnIleArgUlyGlnUlyGlnAlaLeu 343
DB 307 GCGGATGTCACAGCTGCCGGAANAAAGATCACCGCTGCGAGAGAAATTAGAGAGATG 366
QY 344 ThrGlnUlySerGlnUlyValProIleGlnProAspValGlnThrGlnLeuLeuAspArg 363
DB 367 GCGAGCGCGGAGCAAGTCCCATCGAGCCGAGAGGCGAGACAGACATCTCGAGCGCTCTC 426
QY 364 GlnGlnUlyLeuProGlyCysValGlyGlyValValProGlyValGlyUlyTyrAspAla 383
DB 427 AGCAGCGTCMAAAGGCTATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 486
QY 384 AlaValLeuValLeuGlnUlyAsnGln 391
DB 487 GCTCTCTACTGAAAGCAGATGAG 510
```

RESULT 29

ABL60246 standard; DNA; 728 BP.

```
XX ABL60246;
XX
DT 29-JUL-2002 (first entry)
XX Gossypium hirsutum PMVK encoding DNA SEQ ID NO 4.
DE
XX
```

KM Upland cotton; PMVK; phosphomevalonate kinase; plant; herbicide;
KM growth regulator; ds.

XX Gossypium hirsutum.

XX DE10057755-A1.

XX 23-MAY-2002.

XX 22-NOV-2000; 2000DE-1057755.

XX 22-NOV-2000; 2000DE-1057755.

XX 22-NOV-2000; 2000DE-1057755.

XX (FARB) BAYER AG.

XX Meissner R, Lechelt-Kunze C;

XX WPI; 2002-445360/48.

XX New nucleic acid encoding plant phosphomevalonate kinase, useful for

XX identifying modulators, potentially useful as herbicides and growth

XX regulators -

XX Disclosure; Page 14-15; 18pp; German.

XX The invention relates to a nucleic acid (I, ABL60244) that encodes a

XX plant phosphomevalonate kinase (PMVK, ABB7505), excluding the known

XX fully defined partial sequences ABL60245-ABL60247. Plant PMVK (I),

XX constructs and host cells that contain (I) are used to identify agents

XX that bind to and/or modulate activity of PMVK, potentially useful as

XX herbicides and growth regulators. (I) is also used for recombinant

XX production of PMVK. The present sequence is that of a PMVK encoding DNA

XX excluded by the invention.

XX Sequence 728 BP; 213 A; 151 C; 144 G; 219 T; 1 other;

Alignment Scores:

Pred. No.:	2.64e-14	Length:	728
Score:	231.00	Matches:	76
Percent Similarity:	43.63%	Conservative:	37
Best Local Similarity:	29.34%	Mismatches:	60
Query Match:	10.36%	Indels:	86
DB:	24	Gaps:	11

US-10-069-062-7 (1-432) x ABL60246 (1-728)

```
QY 17 TTTLeuValLeuGlnProIleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHis 36
DB 7 TACGCTATTGTTAAGCAATTGATGAGCT----- 36
QY 37 AlaValIleThrProUlyGlyThrSerLeuUlyGlnSerArgIleUlyHisSerSerPro 56
DB 37 -----ATCAAGCTCGAAAGCGCGGCGATGCTTGGACCGCATGTCAAGTCACTACTCT 90
QY 57 GlnPheAlaAsnGlyGlnUlyTrpGlnUlyHisSer----- 68
DB 91 CAGCTTCCAGAG---GAAGCATGTATAATGTCTCGGAACATTAAACACTTCAGTCT 147
QY 69 ---SerAsnThrGlnUlyProArgGlnUlyAlaGlnSerArgIleAsnProPheLeuGlnUly 87
DB 148 GATCTTCMAAGTGA---TCAGG---AACCTTTGTGAAAAAT 186
QY 88 ThrIlePheIleValLeuAlaTyrIleGlnProThr-----GlnAla 101
DB 187 GCTATTCAATATATACATATAGACAGCTGACATGACATTTGACAGAAATPAGAAAGAGCA 246
QY 102 PheAsp-----LeuGlnUlyIleIleTyrSerAspProGlyTyrHis 115
DB 247 TTAGATTAACACTACTTACAAAGCTTGTATATATAGATCTTATGATGATGATGATGAT 306
QY 116 SerGlnUlyAsp----- 119
DB 307 TCATACAGGAATCAGATAGAGACACTGTGCTTCGTTGACACACTGAAGCAATTGGCTACT 366
```

QY 120 -----ThrglurhlyserSerSeranGlyGluysThrphe 132
DB 367 CTACCAACGCTTACATCAATTCATTCATTCGAGCAATCAATGGA----- 414
QY 133 LeuYrHISerArAlaIleThrgIuValGluysThrgIySerSerAlaGly 152
DB 415 -----GCAATTGCAACCTGAAAGTGAAGAAACGAGATGGGTTTCATTCGACGA 465
QY 153 LeuValSerValValAlaIleThrSerLeuSerSerHisPheIle----- 166
DB 466 ATGACAACTGCTGTACTGCTGCTTACTT---CATTAATCTGGTGTGTAACCTTCC 522
QY 167 -----ProanValIleSerThraAsnlysaPheIleuHis 178
DB 523 ACCCTTTTCGACATCAACACCAAGAAATTAAGAAATTCACAGATCTCATTTGTCAT 582
QY 179 AnnValAlaGlnIleAlaHisCySerYrAlaGlnlyValIleGlySerGlyPheAspVal 198
DB 583 ATGATATGCTCAAGTCCCACTGATATGCCCCAAGGTAAAGTTGGCGATGCTTGTATGTC 642
QY 199 AlaThrAlaIleTyrGlyLeuIleValTyrArgArgPheGlnProAlaIleuLeasn 217
DB 643 AGTTCTGCTGTATGGAGTCAAGCGTTATGTTTCGTTTCAACAAAGTGTCTTCT 699
RESULT 30
ABT19456
ID ABT19456 standard; DNA; 5158 BP.
XX
AC ABT19456;
XX
DT 16-APR-2003 (fixet entry)
XX
DE Aspergillus fumigatus essential gene #1814.
XX
KM Fungicide; cytosstatic; essential gene; Aspergillus fumigatus; infection;
XX
KM cancer; contamination; biotilm; antibody; immune response; de.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PE 23-APR-2002; 2002WO-US13142.
XX
PR 23-APR-2001; 2001US-285697P.
XX
PR 27-APR-2001; 2001US-287066P.
XX
PR 05-JUN-2001; 2001US-295890P.
XX
PR 09-JUL-2001; 2001US-303899P.
XX
PR 31-AUG-2001; 2001US-316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of
XX
PT Aspergillus fumigatus, useful for treating or preventing infections by
XX
PT A. fumigatus, or for treating a non-infectious disease in a subject
XX
PT e.g. cancer -
XX
PS Disclosure; Page -; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
XX
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
XX
CC the invention are used to treat or prevent infections by a pathogenic
XX
CC organism such as A. fumigatus, to treat a non-infectious disease in a
XX
CC subject (e.g. cancer), to prevent or contain contamination of an object
XX
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
XX
CC biotilm comprising A. fumigatus. The polynucleotides are useful for
XX
CC expressing recombinant protein for characterisation, screening or

CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case of virulence
CC factors. This polynucleotide sequence represents one of the essential
CC genes of Aspergillus fumigatus of the invention.

SQ Sequence 5158 BP; 1199 A; 1399 C; 1300 G; 1260 T; 0 other;

Alignment Scores:
Pred. No.: 3,096-11 Length: 5158
Score: 213.00 Matches: 61
Percent Similarity: 38.91% Conservative: 25
Best Local Similarity: 27.60% Mismatches: 69
Query Match: 9.55% Indels: 66
DB: 25 Gaps: 5

US-10-069-062-7 (1-432) x ABT19456 (1-5158)

QY 4 AlaPheSerAlaProGlyValAlaPheLeuAlaGlyIleYrLeuValLeuGluProIle 23
DB 4490 GCCTTCTCCGCGCGGGAAGTCTCTCTCACTGGGGTTATCTGTCGACCGCAGC 4549
QY 24 TyrAspAlaYrValThraIleuSerSerArgPheHisAlaValIle----- 39
DB 4550 TACACTGGAGCTGTGTGCTGCTTGCACCGCAATCATCATGTTTCAGCAATTGAGA 4609
QY 40 -----ThrProlys 42
DB 4610 CGGAACATCGCGGGAAGCCGCTCGGATCCGCGATGCCGCGGTGCGACACTTCG 4669
QY 43 -----GlyThrSerLeuysGlnSerArgIleYleIleSerSerPro 56
DB 4670 GCGGAAGCATCTTTCATGAGACAGAGAAGACAGACAGATGCTGTACTCTCCCA 4729
QY 57 GlnPheAlaenglyGluTrpGluYrHisIle----- 67
DB 4730 CAGTTCTGATGCGGATGAGATATGATATACAAAGATCGAGACGCTGAGAGATC 4789
QY 67 ----- 67
DB 4790 CTAGTAAACAGAGAAATGATGGGTAAAGCAACATATTATAGTACGTAATTGCCA 4849
QY 68 SerSerAenThrGluysProArgGluValGlnSerArgIleAsnProPheLeuGluAla 87
DB 4850 GAGCTGACCATTCAGGCCACGC-----AACCTTTGTGCAGACC 4891
QY 88 ThrIlePheIleValLeuAlaYrIle-----GlnProThrGluAlaPheAsp 103
DB 4892 TCTCGAACTTCGCTTACCTTACATCATCATGATGCGGCACTCGAAGATTCGGGCA 4951
QY 104 LeuGluIleIleIleYrSerAspProGlyYrHisSerGlnGluAspThrGluTrlys 123
DB 4952 TTATCGATTACTATCTCCGCGACAGATTAATCTCGACGCGCTTCTCCAGAGCT 5011
QY 124 ThrSerSerAnGlyGluYrThraPheLeuYrHisSerArgAlaIleThrgIuValGlu 143
DB 5012 TCGGACCTCCGGTGTCAAGAGATTCGTGATCTTGTGTCCTTCAGAGGACAC 5071
QY 144 LysThrgIleuGlySerSerAlaGlyLeuValSerValAlaIleThrSerLeuSer 163

Db	5072	AAGACAGCGCTGGCGCTTTTCAGCGCCGCTTGTCATCGCCGCTGATCTCTGTATC	513
QY	164	His 164	
Db	5132	CAC 5134	
RESULT 31			
XX	AAAF27596	AAAF27596	
ID	AAAF27596	standard; DNA; 1107 BP.	
XX	AAAF27596;		
XX			
DT	30-MAR-2001	(first entry)	
XX			
DE	Mevalonate pathway gene #3.		
XX			
KW	Mevalonate pathway; disease; infection; ds.		
XX			
OS	Enterococcus faecalis.		
XX			
FN	WO200078935-A1.		
XX			
PD	28-DEC-2000.		
XX			
PE	22-JUN-2000; 2000MO-US17262.		
FR	22-JUN-1999; 99US-0140519.		
PR	02-AUG-1999; 99US-0146682.		
XX			
PA	(SMIK) SMITHKLINE BEECHAM CORP.		
XX	(SMIK) SMITHKLINE BEECHAM PLC.		
PI	Brown JR, Gwynn M, Matchie TB, Myers JE, Traini CM, Van Horn S;		
PI	Wilding EJ;		
DR	WPI; 2001-071392/08.		
XX			
PT	New isolated mevalonate pathway gene polynucleotide derived from		
PT	bacterium is useful for treatment of bacterial infection -		
PS	Claim 20; Page 13; 158bp; English.		
XX			
CC	The present invention relates to an isolated mevalonate pathway gene		
CC	derived from a bacterium from clade of class II of the		
CC	phylogenetic tree referred to in the specification. The invention may be		
CC	used for treatment of disease related to bacterial infection, e.g.		
CC	conjunctivitis, pneumonia, bacteraemia and meningitis.		
XX			
SO	Sequence 1107 BP; 347 A; 200 C; 240 G; 320 T; 0 other;		
Alignment Scores:			
QY	Pred. No.: 7.99e-12	Length: 1107	
	Score: 209.50	Matches: 113	
	Percent Similarity: 39.38%	Conservative: 52	
	Best Local Similarity: 26.97%	Mismatches: 137	
	Query Match: 9.39%	Indels: 117	
DB:	22	Gaps: 23	
US-10-069-062-7 (1-432) x AAF27596 (1-1107)			
QY	6	SeralAPROGILyVAlaPheLeuAlaGlyLYrLeuValLeuGluProIleTyraSP	25
	:::		
Db	13	ACTAGCCAGGAAAGTTATTATTCACAGAGATATGCCGTTTGAACTGGCCACCT	72
	:::		
QY	26	AlaTYValThAlaLeuSerSerArgmethIAlaValIlethbProlyseGlyThSer	45
Db	73	GCCATTATCGTTCGTGTGAT-----CAATTCTGTAAGTGTACTGTGAAAGAAACAACA	126
	:::		
QY	46	LeuYsgLSeuSerArgIleLysIleSerSerProGlnPheAla-----	59
	:::		
Db	127	GATGAAGCAAGT-----ATTCAATCTGCACAAATACAGTCTTTACTGATTGTTGG	177
	60	-----AaenGlyLutPrluTyHisIleSerSerAsnThnGluysProlArgGlu	76

Db	178	ACAGCCGAATGGAGCTCGTATTAGATATT-----	210
Qy	77	ValGInSerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrlle	96
Db	211	-----CGCGAAATATCCTTT-----CATTANTGTTTACGGCGGAT	246
Qy	97	GInProThnGluAlaPheAspLeuIuIleIleIleTySerAspProIuTyRhIser	116
Db	247	CATTAACCTGAAAAATAT-----GGC	267
Qy	117	GInGluAepThnGluThrTySerSerSerAengIuGluSerThrPheLeuTyRhIser	136
Db	268	CAAGG-----CAAAACAAGATTCGATTT---TATCATTTA	303
Qy	137	ArgAlaIleThnGluValGlu-----LyeThnGlyLeuGlySerSerAla	151
Db	304	AAAGTACAGAGTATGATAGTATGTTCAAAATGACAGCAAAATATGCTTGTTCAAC---	360
Qy	152	GlyLeuValSerValAlaAlaThrsereLeuLeuSerRhIspheIleProAsnValIleser	171
Db	361	GGTGACAGTAAACCGTTGGAACGTGCAAAGCCCTGAATATTTTATGAC-----TTAGGT	414
Qy	172	ThrAsnIuAspIleLeuHisAsnValAglInIleAlaHisCyetyAlaGlnIuSlys	191
Db	415	TTGCAAAATGAGAAATTTTCAATATTAACAGATTAAGTCACTTACCGCTTCAAGAAAT	474
Qy	192	IleGlySerGlyPheAspValAlaThrAlaIetyr---GlyLeuIleValTyRArgArg	210
Db	475	---GGTTCTTGCGAGATATGCGCGCAGCTGTATCGGGGGCTGAGTTCCTTTCAAC	511
Qy	211	PheGInProAlaLeuIleAsnAspValPheGInValLeuGInSerAspProGluTyRpe	230
Db	532	TTTCATCATCATGTTGGTCAAT-----CAAAAAGTA	561
Qy	231	ProThnGlu---LeuTyLeuTyLeuIleGInSerAsnTrpGluGlu-----	244
Db	562	ACCACTGAAACCTTAACGATTTGTTAGCAATGACGTGCGCTGAATTAATGATTTTCCG	621
Qy	245	---LysHISgluArgCysThrLeuProTyrglyIleLysLeuLeuMetGlyAspValLys	263
Db	622	TTAANA-----GTACCGAAACAATACGTTTCTAATGATTGTTGACAGGT	666
Qy	264	GlyGlySerGluThnProLysLeuValSerArgValLeuGInTrpLysGlySerGluTyRPro	283
Db	667	AGTCTGTGCGTCACTTCAAGACTTAGTATGATCGAGTTCAATCAAAAGAGAAAA---	723
Qy	284	GluGInSerSerValTyRAspGInLeuAsnSerAlaAsnLeuGInPheMetLysGlu	303
Db	724	-----CAAGCGGCTTATGACGACGTTTCTTAATGAAA	753
Qy	304	LeuArgGluMetArgGluLysTyRAspSerAspProGluThnTyrlleLysGluLeuAsp	323
Db	754	AGTGGCTTGT-----GTGCAAAACAATGATTAAT-----GGC	786
Qy	324	HisSerValGluProLeuThnValAlaIleLysAsnIleArgLysGlyLeuGInAlaLeu	343
Db	787	TTTAAACAAGAAAAATTTCTGTATTCTCAAAAAACAATTAACATAAATGCGCAATTCCTC	846
Qy	344	ThnGInLysSerGlu-----ValProIleGluProAspValGInThnGInLeuLeu	360
Db	847	GCCGATTAATCTCACTGACTGCTGTGTGTAATCGAAACAGAGCCTTGAAAAATCTTGT	906
Qy	361	AspArgCysGInGluIleProGlyCysValGlyGlyValValProGlyValAglGlyTyR	380
Db	907	GATTTGGCTGAATCTTATACAGAGCTCGAAATCT-----TCTGGCGCTGGCGGGGCG	960
Qy	381	AspAlaIleAlaValLeuValLeuGluAsnGlnValGlyAsnPheLysGlnTyR	399
Db	961	GATGTGTGGATTTGTATTT-----TTCCGCCAAAAATCT	993

ID	Accession	Standard	DNA	1107 BP
XX	AA552922	standard	DNA	1107 BP
XX	AA552922			
AC	13-FEB-2002	(first entry)		
DT	13-FEB-2002	(first entry)		
XX	Enterococcus faecalis	DNA for cellular proliferation protein #350.		
DE	Enterococcus faecalis	DNA for cellular proliferation protein #350.		
XX	Antisense; ds; prokaryotic cellular proliferation gene;			
KM	antibiotic; antibacterial; drug design.			
XX	Enterococcus faecalis.			
OS	MO200170955-A2.			
PN	27-SEP-2001.			
XX	27-SEP-2001.			
PD	21-MAR-2001; 2001WO-US09180.			
XX	21-MAR-2001; 2001WO-US09180.			
PR	21-MAR-2000; 2000US-191078P.			
PR	23-MAY-2000; 2000US-206848P.			
PR	26-MAY-2000; 2000US-207777P.			
PR	23-OCT-2000; 2000US-242578P.			
PR	27-NOV-2000; 2000US-233625P.			
PR	22-DEC-2000; 2000US-257931P.			
PR	16-FEB-2001; 2001US-269308P.			
XX	(ELIT-) ELITRA PHARM INC.			
PA	Haselebeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;			
XX	Yamamoto RT, Xu HH;			
PI	WPI; 2001-611495/70.			
DR	P-PSDB; AAU35063.			
XX	New polynucleotides for the identification and development of			
PT	antibiotics, comprise sequences of antisense nucleic acids -			
XX	Claim 27; Seq ID No 6559; 511pp; English.			
PS	The invention relates to antisense inhibitors of genes essential to			
XX	prokaryotic cellular proliferation, their use in identifying the			
CC	genes, their use in the discovery of novel antibiotics, the essential			
CC	genes themselves and the encoded proteins. The prokaryotes used are			
CC	Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella			
CC	pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The			
CC	invention is also useful for the identification of potential new targets			
CC	for antibiotic development. The antisense nucleic acids can also be used			
CC	to identify proteins used in proliferation, to express these proteins,			
CC	and to obtain antibodies capable of binding to the expressed proteins.			
CC	The proteins can be used to screen compounds in rational drug discovery			
CC	programmes. The antisense nucleic acid sequence is also useful to screen			
CC	for homologous nucleic acids which are required for cell proliferation in			
CC	a wide variety of organisms. The present sequence encodes an			
CC	essential prokaryotic cellular proliferation protein.			
CC	Note: The sequence data for this patent did not form part			
CC	of the printed specification, but was obtained in electronic			
CC	format directly from WIPO at			
CC	ftp.wipo.int/pub/published_pct_sequences.			
XX	Sequence 1107 BP; 347 A; 200 C; 240 G; 320 T; 0 other;			
XX	Alignment Scores:			
XX	Pred. No.:	7.99e-12	Length:	1107
XX	Score:	209.50	Matches:	113
XX	Percent Similarity:	39.38%	Conservative:	52
XX	Best Local Similarity:	26.97%	Mismatches:	137
XX	Query Match:	9.39%	Indels:	117
XX	DB:	23	Gaps:	23
XX	US-10-069-062-7 (1-432) x AA552922 (1-1107)			

[illegible]

QY 344 ThrGlnLysSerGlu-----ValProIleGluProAspValGlnThrGlnLeu 360
 Db 3855 GCGGATTTATCTTCACTGACTGCTGTGTGTATGAGAAACGAACTTGAATAATCTTTGT 3914
 QY 361 AspArgCysGlnGlnIleProGlyCysValGlyValValProGlyValGlyGlyTyr 380
 Db 3915 GATTGTGCTGAATCTTATACAGAGCTGCGAATCT-----TCTGCGCTGGCGGGGCG 3968
 QY 381 AspAlaIleAlaValLeuValLeuGlnValGlyAsnPhelGlnGlyThr 399
 Db 3969 GATTGTGGAATGTATAT-----TTCCGCAAAATCT 4001
 RESULT 34
 ABS98936
 ID ABS98936 standard; DNA; 6835 BP.
 XX ABS98936;
 AC
 XX
 DT 18-DEC-2002 (first entry)
 DE Enterococcus faecalis contig sequence #204.
 XX
 XX Computer readable medium; Enterococcus faecalis; microbe; growth;
 KM pathogenicity; vaccine; resistance; Enterococcal infection; commercial;
 KM therapeutic; industrial; fermenting; sugar source; metabolite; vaccine;
 KM biotech technology; antibacterial; modulator of nucleic acid expression;
 KM contig; ds.
 OS Enterococcus faecalis.
 XX
 XX US2002120116-A1.
 PN
 XX 29-AUG-2002.
 PD
 XX 04-MAY-1998; 98US-0070927.
 PF
 XX 04-MAY-1998; 98US-0070927.
 PR
 XX 04-MAY-1998; 98US-0070927.
 XX
 PA (KUNS/) KUNSCH C A.
 PA (DILL/) DILLON P J.
 PA (BARA/) BARASH S.
 XX
 PI Kunsch CA, Dillon PJ, Barash S;
 DR MPI; 2002-750065/81.
 XX
 XX Computer readable medium having recorded on it a Enterococcus faecalis
 PT nucleotide sequence useful for detecting diseases related to
 PT Enterococcus infections in animals -
 XX
 PS Claim 1; Page -; 11pp; English.
 CC The present invention relates to a new computer readable medium with an
 CC Enterococcus faecalis nucleotide sequence. The invention is useful to
 CC diagnose the presence of E. faecalis in a sample or determining the
 CC presence of a specific microbe in a sample. The invention is also useful
 CC for modulating the growth or pathogenicity of E. faecalis, in a vaccine
 CC to confer resistance to Enterococcal infection, for commercial,
 CC therapeutic and industrial purposes, and for fermenting a particular
 CC sugar source or to produce a particular metabolite. The invention is
 CC useful for detecting diseases related to Enterococcus infections in
 CC animals, and for detecting E. faecalis using biochip technology. The
 CC present nucleic acid sequence represents an Enterococcus faecalis contig
 CC DNA sequence of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification but was obtained in electronic format directly
 CC from USPTO at <http://seqdata.uspto.gov>.
 XX
 SO Sequence 6835 BP; 2065 A; 1291 C; 1370 G; 2102 T; 7 other;
 Alignment Scores: 1.06e-10 Length: 6835
 Pred. No.:

Score: 209.50 Matches: 113
 Percent Similarity: 39.38% Conservative: 52
 Best Local Similarity: 26.97% Mismatches: 137
 Query Match: 9.39% Indels: 117
 DB: 24 Gaps: 23
 US-10-069-062-7 (1-432) x ABS98936 (1-6835)
 QY 6 SerAlaProGlyValAlaPheLeuAlaGlyTyrLeuValLeuGluProIleTyrAsp 25
 Db 3021 ACTACGCCAGAAAGTATTTATATGACAGAAATAGCCGTTGTCAGCCGCACT 3080
 QY 26 AlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLysGlyThrSer 45
 Db 3081 GCCATTATCGTTCGTGTGAT-----CAATTGTAACCTGTAACCTGTAACCAACA 3134
 QY 46 LeuLysGluSerArgIleLysIleSerSerProGlnPheAla----- 59
 Db 3135 GATGAGGCGAGT-----ATTCAATCTGCACAAATACAGCTTTACTTATTCGTTGG 3185
 QY 60 -----AsnGlyLutProGlyTyrHisIleSerSerAsnThrGlyLeuProArgGlu 76
 Db 3186 ACACGCCCAAAATGTGAGCTGTATATAGATAT----- 3218
 QY 77 ValGlnSerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIle 96
 Db 3219 -----CGGAAATCTTT-----CATATGTTTACGCGGAT 3254
 QY 97 GlnProThrGluAlaPheAspLeuGlnIleIleIleTyrSerAspProGlyTyrHisSer 116
 Db 3255 CATTAACCTGAATAATAT-----CGG 3275
 QY 117 GlnGluAspThrGlnThrLysThrSerSerAsnGlyLysThrPheLeuTyrHisSer 136
 Db 3276 CAAGAG-----CAAAACAAGAAATGCTATTT-----TATCATTTA 3311
 QY 137 ArgAlaIleThrGluValGlu-----LysThrGlyLeuGlySerSerAla 151
 Db 3312 AAAGTACAGAGTGAATTTAGATATGTTCAAAATGACGAAATATGCTTGTTCAGC--- 3368
 QY 152 GlyLeuValSerValValAlaThrSerLeuSerHisPheIleProAsnValIleSer 171
 Db 3369 GGTGCGTAAACCGTTGGAACGTCAAGCCTTGAATATTTTATGAC-----TTAGCT 3422
 QY 172 ThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAlaGlnLys 191
 Db 3423 TTGAAAATGAGAAATTTTCAATTTATACCATTAAGCTTAAACCGTTCAAGAAAT 3482
 QY 192 IleGlySerGlyPheAspValAlaThrAlaIleTyr--GlyLeuIleValTyrArgArg 210
 Db 3483 ---GGTTCTTCCGAGATATGCGCCGACGCTGTTACGGGGGCTGGATTGCTTTCAAC 3539
 QY 211 PheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGluLysPhe 230
 Db 3540 TTCGATCATGATGGGTCAAT-----CAAAAGTA 3569
 QY 231 ProThrGlu--LeuLysLysLeuIleGluSerAsnTyrGluGlu----- 244
 Db 3570 GCCATGAACATTAAGTATGATTTGTTCATGACATGACGCTGTAATGATTTTCCG 3629
 QY 245 ---LysHisGluAsnGlyThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLys 263
 Db 3630 TTAAAH-----GTACCAAAACACTACCTTACTAATGATTTGGACAGCT 3674
 QY 264 GlyLysSerGlnThrProLysLeuValSerArgValLeuGlnTyrLysGlnLysPro 283
 Db 3675 AGTCCTGGTCACTTGCAGATTGATGATGAGTCAATCAATCAAAAGAAAGAAA--- 3731
 QY 284 GluLysSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGlu 303
 Db 3732 -----CAACGCGCTTATGACAGCTTCTTATGAAA 3761
 QY 304 LeuArgLysMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAsp 323

Db 3762 AGTCGCTTGT-----GTGGAACATGATTAAT-----GGC 3794
 QY 324 HisSerValGluProLeuThrValAlaIleValAsnIleArgIleGluGlnAlaLeu 343
 Db 3795 TTTTACACAGAAAAATTTCTGTATTCAAAAACAATTAATAAATGCCCAATGCTC 3854
 QY 344 ThrGlnIleSerGlu-----ValProIleGluProAspValGlnThrGlnLeu 360
 Db 3855 GCCGAATTATCTTCACTGAGCTGTGTGTATCGAAGAACAGACCTTGAAAAATCTTTGT 3914
 QY 361 AspArgCysGlnGluIleProGlyCysValGlyValValProGlyAlaGlyIleTyr 380
 Db 3915 GATTTCGCTGAATCTTATACAGAGCTGCCAATCT-----TCTGCGCTGCGCGGCGG 3968
 QY 381 AspAlaIleAlaValLeuValIleGluAsnGlnValGlyAsnPhelSerGlnIleTyr 399
 Db 3969 GATTGTGGATTTGTAAT-----TTCCGCCAAAAATCT 4001

RESULT 35

AAAF27619 standard; DNA; 1077 BP.

AAAF27619;

30-MAR-2001 (first entry)

Mevalonate pathway gene #5.

Mevalonate pathway; disease; infection; da.

Staphylococcus haemolyticus.

MO200078935-A1.

28-DEC-2000.

22-JUN-2000; 2000MO-US17262.

22-JUN-1999; 99US-0140519.

02-AUG-1999; 99US-0146682.

(SMIK) SMITHKLINE BEECHAM CORP.

(SMIK) SMITHKLINE BEECHAM PLC.

Brown JR, Gwynn M, Mathie TB, Myers JE, Traini CM, Van Horn S;

Wilding EJ;

WPI; 2001-071392/08.

New isolated mevalonate pathway gene polynucleotide derived from

bacterium is useful for treatment of bacterial infection -

Claim 20; Page 33; 158pp; English.

The present invention relates to an isolated mevalonate pathway gene

derived from a bacterium from clade of Class II of the

CC phylogenetic tree referred to in the specification. The invention may be

used for treatment of disease related to bacterial infection, e.g.

conjunctionitis, pneumonia, bacteremia and meningitis.

Sequence 1077 BP; 357 A; 163 C; 220 G; 337 T; 0 other;

Alignment Scores:

Pred. No.: 3.91e-10 Length: 1077
 Score: 193.00 Matches: 101
 Percent Similarity: 38.95% Conservative: 63
 Best Local Similarity: 23.99% Mismatches: 129
 Query Match: 8.65% Indels: 128
 DB: 22 Gaps: 21

US-10-069-062-7 (1-432) x AAFA27619 (1-1077)

QY 7 AlaProGlyIleValAlaPheLeuAlaGlyIleTyrLeuValIleGluProIleTyr----- 24
 Db 16 GCACCAAGAAAGCTTATAGTACAGTGAATATGACGTGACGACGATATAGTCT 75
 QY 25 -----AspIleTyrValThrAlaLeuSerSerArgMechIleValIle 39
 Db 76 GTCTTAATTCGCGTGCATGATTTGTTCACGCTTCATTAATGAAGCTTCATACAGTAACA 135
 QY 40 Thr-----ProIleGlyThrSerLeuIleGluSer 49
 Db 136 AGTACGATTCATCCACAGACATTATGACCTGTAACCTTTAATCGCATGACATGACAT 195
 QY 50 ArgIleIleIleSerSerProGlnPheAlaGlnGlyIleTyrGluTyrHisIleSerSer 69
 Db 196 AAAATTGATATCTCCGATGCTAATGCTGATACAA----- 231
 QY 70 AsnThrGluIleProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThrIle 89
 Db 232 -----TyrAAAGTATGTTGA 246
 QY 90 PheIleValIleAlaTyrIleGlnProThrGluAlaPheAspLeuGluIleIleTyr 109
 Db 247 ACTGCAATTGAAGCTTTTCACACATATGCGAAGAGTTGCAAGCTCAATG----- 297
 QY 110 SerAspProGlyTyrHisSerGlnGluIleProThrGluTyrIleTyrSerSerArgGlyIle 129
 Db 298 -----AAGCATTTTCATTTAGAAATCGATAGTAATTTAGATGATGCTCAGCT----- 345
 QY 130 LysThrPheLeuTyrHisSerArgAlaIleThrGluValGluIleTyrThrGlyLeuGlySer 149
 Db 346 -----AATTAATATGCGCTTGCT 366
 QY 150 SerAlaGly-----LeuValSerValAlaThrSerLeuSerHisPheIleProAsn 168
 Db 367 AGTCGCGCAGTTTATGTCACGTCTC-----AAAGCATTAATGAGCTTTATACATATG 420
 QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
 Db 421 CAATATATCAACCTA-----TATATTTATTAATGAGCTGATTTTATATGCGATTA 474
 QY 189 GlnIleLysIleGlySerGlyPheAspValAlaThrAlaIleTyr-----GlyLeuIleVal 207
 Db 475 CAAGTTTAAGCTCATGTGT-----GATATAGCTGTAAAGTATACAGTGTGCTGCTAGCT 531
 QY 208 TyrArgArgPheGlnProAlaLeuIleAsnAsp-----ValPheGlnValIleLeuSerAsp 226
 Db 532 TATAGTACTTTCAT-----CACGATGGGTCAACACACAGATGAGAA----- 576
 QY 227 ProGluIlePheProThrGluLeuLysIleLeuIleGluSerAsnTyrProGluIleLysHis 246
 Db 577 -----ACATCAGTTAATGAGATTAAGAAAAAATTTGGCGGCTCTTCAT 621
 QY 247 ---GluArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValIleGlyIle 265
 Db 622 ATTGAACCTTTACAGCCCCACAGAGATATGAAAGCTTAATGCTGTGACAGCTTCGCT 681
 QY 266 SerGluThrProLysLeuValSerArgValIleGlnTyrIleLysGluIleProGluIle 285
 Db 682 GCTTATACCTCATTTATAGTACGTGAAGTGAAGCCCTTAATA----- 723
 QY 286 SerSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGluLeuArg 305
 Db 723 ----- 723
 QY 306 GluMetArgGluLysTyrAspSerAspProGluIleTyrIleLysGluLeuAspHisSer 325
 Db 724 -----TCAGATCTTCTTTTATGAAAGTTCCTGATCAATCT 762
 QY 326 -----ValGluProLeuThrValAlaIleLys-----AsnIleArgIleGlyLeu 340
 Db 763 CATACATGTTGTGAATAATCTTATCTATGCGTTTAAACAGATTAATAT-----AAAGGTGTT 819
 QY 341 GlnAlaLeuThrGlnLysSerGluValProIleGluPro----- 353


```
DB 820 CCGAATATATTCGACAAATCGTATGATTATTCACAAATGATATGAGGACAGTC 879
QY 354 AspValGlnThrGlnLeuLeuApaCyGsgIngluileProgl1yCyValGlyGlyVal 373
DB 880 GACATTGAAACCGAAATTTAAATAAGTTATGATGATTTGAGAACCTTATGCTGCT 939
QY 374 -----ValProgl1yAglGlyGlyTyraPalaileValleuValleuGln 391
DB 940 GCCAAGACATCAGTCTGCGGTGATTCG---GGAATCGCATTTATGATATTCG 996
QY 392 Val 392
DB 997 ATT 999
RESULT 36
ABQ82712
ID ABQ82712 standard; cDNA, 757 BP.
AC ABQ82712;
XX
XX 03-JAN-2003 (first entry)
XX
XX Rice mevalonate kinase encoding cDNA SEQ ID NO:21.
XX
XX Mevalonate kinase, enzyme; herbicide; agricultural; plant; gene; ss.
XX
XX Oryza sativa.
XX
XX Key Location/Qualifiers
XX CDS 14..550
XX FT /*tag= a
XX FT /partial
XX FT /product= "mevalonate kinase"
XX FT /note= "no start or stop codons given"
XX
XX US2002119546-A1.
XX
XX 29-AUG-2002.
XX
XX 20-JUL-2001; 2001US-0909745.
XX
XX 05-NOV-1998; 98US-107241P.
XX PR 04-NOV-1999; 99US-0433242.
XX
XX (FALC/) FALCO S C.
XX PA (FAMO/) FAMODU O O.
XX
XX Falco SC, Famodu OO;
XX
XX WPI: 2002-731357/79.
XX DR P-PsDB; ABP53755.
XX
XX New mevalonate kinase polypeptide useful for new herbicide discovery
XX PT and design, and for altering levels of mevalonate kinase in transformed
XX PT cells
XX
XX PS Disclosure; Page 25; 39pp; English.
XX
XX The present invention describes an isolated mevalonate kinase polypeptide
XX (I), having a sequence identity of 80 % based on the Clustal method of
XX alignment when compared to corn, rice, soybean and wheat mevalonate
XX kinase. (II) has herbicide and agricultural activities. A nucleic acid
XX (III) encoding (I) is useful for transforming a cell and producing a
XX transgenic plant. A chimeric gene (III) comprising (II) is useful for
XX altering the level of expression of a squalene biosynthetic enzyme in a
XX host cell, where the expression of the chimeric gene results in
XX production of altered levels of mevalonate kinase in the transformed host
XX cell. (III) is also useful for evaluating a compound for its ability to
XX inhibit the activity of a mevalonate kinase. (I) is useful as a target
XX to facilitate the design and/or identification of inhibitors of these
XX enzymes that may be useful as herbicides. Fragments of (II) are useful
XX to create transgenic plants in which the disclosed polypeptides are
```

```
CC present at higher or lower levels than normal or in cell types or
CC developmental stages in which they are not normally found. The nucleic
CC acid fragments are also useful as probes for genetically and physically
CC mapping the genes that they are a portion of and as markers for traits
CC linked to those genes. The present sequence encodes a rice mevalonate
CC kinase protein from the present invention.
XX
XX SQ Sequence 757 BP; 231 A; 141 C; 186 G; 199 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 3,676-09 Length: 757
XX Score: 181.50 Matches: 47
XX Percent Similarity: 45.64% Conservative: 21
XX Best Local Similarity: 31.54% Mismatches: 56
XX Query Match: 8.14% Indels: 25
XX DB: 24 Gaps: 2
XX
XX US-10-069-062-7 (1-432) x ABQ82712 (1-757)
QY 264 GlyGlySerGluThrProlySerLeuValSerArgValLeuGlnTrpLysGlyPro 283
DB 14 GGAGGATCATTCACCTCCATCATGATGATCTGTGAAACAGTGGCAGAACGACACCT 73
QY 284 GluGlySerSerValValTyraPalaileuValSerArgValLeuGlnTrpLysGly 303
DB 74 CAGAAATCCAAAGACAGACATGAGTAATTTGGGATTTGCTAATTCAGTCTGAGAACCA 133
QY 304 LeuArgGluMetArgGluLysTyraPalaileuValSerArgValLeuGlnTrpLysGly 323
DB 134 CTGAGGAACCTTAAACAACTTGTGATGATCACTGGAGAACCTTAAATCTGTTTAA--- 190
QY 324 HisSerValGluProLeuThrVal-----AlaIleLysAsnIleArgLysGly 331
DB 131 CGATCTTATGTCGTCCTACCGTCAGTACGTAAGTACAGACAGTGGCTACCAATCAATCAA 250
QY 332 -----AlaIleLysAsnIleArgLysGly 339
DB 251 GAACTAATTTGATGATCATTCAGTCCGCAAGAGATGCTTCTTGAATTAAGGCTCAT 310
QY 340 LeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspValGlnThrGlnLeu 359
DB 311 ATGCGAGATGGGAGATGACAGCTGTGTTCATTTGAGCCAGAACTCAACAACTCAACTT 370
QY 360 LeuAspArgCyGsgIngluileProgl1yCyValGlyGlyValProgl1yAglGly 379
DB 371 CTGATGCCACTATGAATATGAGAGGTCTTACTAGCTGAGATTTCTGGGCGCGTGGC 430
QY 380 TyraPalaileuValleuValleu 388
DB 431 TTTGATGCAAGTTTCTCAGTATTTTG 457
XX
XX RESULT 37
XX AAH52762
XX ID AAH52762 standard; DNA, 1077 BP.
XX
XX AC AAH52762;
XX
XX 03-SEP-2001 (first entry)
XX
XX DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:917.
XX
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX KW vaccination; endocarditis; ds.
XX
XX OS Staphylococcus epidermidis.
XX
XX PN MO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX PD 09-NOV-2000; 2000WO-US30782.
XX
XX PF 09-NOV-1999; 99US-0164258.
XX
XX PR
```

XX (GLAXO) GLAXO GROUP LTD.
 PA
 XX
 PI Kimerly WJ;
 DR WPI; 2001-316495/33.
 XX P-PSDB; AAG81912.
 XX
 PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
 useful for vaccinating against infections, e.g. endocarditis -
 XX
 PS Claim 8; Page 272-273; 2188pp; English.
 XX
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce host cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX
 CC
 SQ Sequence 1077 BP; 368 A; 158 C; 219 G; 332 T; 0 other;
 Alignment Scores:
 Pred. No.: 6.05e-09 Length: 1077
 Score: 181.50 Matches: 99
 Percent Similarity: 36.74% Conservative: 52
 Best Local Similarity: 24.09% Mismatches: 131
 Query Match: 8.14% Indels: 129
 DB: Gaps: 18
 US-10-069-062-7 (1-432) x AAH52762 (1-1077)
 QY 7 AlaProGlyLysAlaPheLeuValGlyGlyTyrLeuValLeuGluProIleTyrAspAla 26
 DB 16 GCGCCCGGAAACTTATATATGCGAGCGATATGCAATGCAACCGAGATATAATCT 75
 QY 27 TyrValThrAlaLeuSerSerArgMetHisAlaValIle----- 39
 DB 76 ATCTTATATGCAAGTAAATCGCTTTGTAACGCGCAATGAGCGCTCAATAAAGTTGAA 135
 QY 40 -----ThrProLysGlyThrSerLeuLysGluSer 49
 DB 136 GGTAGATTCATTCCAAAACATTACATTATGAAACGACTTAATGACCTATGAAAGT 195
 QY 50 ArgIleLysIleSerSerProGlnPheAlaAsnGlyGluTyrGluTyrHisIleSerSer 69
 DB 196 AGAATTGAATCTCAAGATGTTCAAGCTGCTAAGCAA----- 231
 QY 70 AsnThrGluLysProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThrIle 89
 DB 232 -----CTGAAATATGTTGTG 246
 QY 90 PheIleValLeuAlaTyrIleGlnProThrGluAlaPheAspLeuIleIleIleTyr 109
 DB 247 ACAGCATATGAAGTGTGAAACGATATGCGCAGATTGCAATGAAATTTA----- 297
 QY 110 SerAspProGlyTyrHisSerGlnGluAspThrGluThrLysThrSerSerAsnGlyGlu 129
 DB 298 -----AAGCACTTCAATTAAACCAATGATGATTAACAGATTAACATCTGCT----- 345

QY 130 LysThrPheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySer 149
 DB 346 -----CAGAGTACGATTAAGTTCA 366
 QY 150 SerAlaGly--LeuValSerValAlaThrSerLeuSerHisPheIleProAsn 168
 DB 367 AGCGCCGCGTGTGTAGTATCTGTGTT-----AAGCTTTGAATGAATTCATATGCT 417
 QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
 DB 418 ---TTGGAATTATCAAACTTATATTAATAATTAGCTGTAATGCAAAATGAAATTA 474
 QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyr---GlyLeuIleVal 207
 DB 475 CAAGATTATAGTTCATGTCGCG--GATATGCGGTAGTACGTTCACAGTGGCTGCA 531
 QY 208 TyrArgArgPheGlnProAlaLeuIleAsnAsp---ValPheGlnValLeuGluSerAsp 226
 DB 532 TATAGTACGTTGAC-----CATGACTGGGTGAACACAGCAATGAGAAAGAA--- 576
 QY 227 ProGluLysPheProThrGluLeuLysLysLeuIleGluSerAsnThrProLysHis 246
 DB 577 -----ACATCGGTGATGATGATGTTTGGAAAAAAATGGCCAGCTTACAT 621
 QY 247 ---GluArgCysThrLeuProTyrGlyIleLysLeuMetGlyAspValLysGlyGly 265
 DB 622 ATCGAACCTTTCAAGCTCCGAAATATGAAAGTCTTATGATGATGACGTGGTCTCA 681
 QY 266 SerGluThrProLysLeuValSerArgValLeuGlnTyrLysGluLysProGluGlu 285
 DB 682 GCTTCTTCCACACTGATGATGATGAGTCAAACTTTAAAA----- 723
 QY 286 SerSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGluLeuArg 305
 DB 723 ----- 723
 QY 306 GluMetArgLysLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHisSer 325
 DB 724 -----TCAGATCCAAGTTTATGATGATGATTTTATATCATCT 762
 QY 326 -----ValGluProLeuThrValAlaIleLys---AsnIle----- 336
 DB 763 CATGCTTGTTGAAAGTATTAATCCAGCTTTAAACTTAATAATATCAAAAGGTTCAA 822
 QY 337 -----ArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValPro 350
 DB 822 AAGATGATACGTATTAACAGAGATTTATTCATATCTATGATGATACGAAGCATCAGTTGAA 882
 QY 351 IleGluProAspValGlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysVal 370
 DB 883 ATTTGAACAGATTAAGCTAAAAAATTATGTATGATGTCGGTGAAGAACGCGTGGC----- 936
 QY 371 GlyGlyValValProGlyValAlaGlyTyrTyrAsp 381
 DB 937 GCTTCTAAAACTTCAAGTGTGCTGTGCGTGCAT 969
 RESULT 38
 AAF27620
 ID AAF27620 standard; DNA; 1077 BP.
 XX
 AC AAF27620;
 XX
 DT 30-MAR-2001 (first entry)
 XX
 DE Mevalonate pathway gene #5.
 XX
 KM Mevalonate pathway; disease; infection; ds.
 XX
 OS *Staphylococcus epidermidis*.
 XX
 PN WO200078935-A1.
 XX

PD 28-DEC-2000.
 XX 22-JUN-2000; 2000WO-US17262.
 PF 22-JUN-1999; 99US-0140519.
 PR 02-AUG-1999; 99US-0146682.
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX Brown JR, Gwynn M, Machie TB, Myers JE, Traini CM, Van Horn S,
 PI Wilding BJ;
 XX WPI; 2001-071392/08.
 DR New isolated mevalonate pathway gene polynucleotide derived from
 PT bacterium is useful for treatment of bacterial infection -
 XX
 XX Claim 20; Page 34; 158pp; English.
 PS The present invention relates to an isolated mevalonate pathway gene
 CC derived from a bacterium from clade of Class II of the
 CC phylogenetic tree referred to in the specification. The invention may be
 CC used for treatment of disease related to bacterial infection, e.g.
 CC conjunctivitis, pneumonia, bacteraemia and meningitis.
 CC
 XX Sequence 1077 BP; 370 A; 159 C; 217 G; 331 T; 0 other;
 SQ

Alignment Scores:
 Pred. No.: 6.05e-09 Length: 1077
 Score: 181.50 Matches: 99
 Percent Similarity: 36.74% Conservative: 52
 Best Local Similarity: 24.09% Mismatches: 131
 Query Match: 8.14% Indels: 129
 DB: 22 Gaps: 18

US-10-069-062-7 (1-432) x AAF27620 (1-1077)

QY 7 AAlAPROGLYLYSALPheleuAaGlyGlyTyrLeuValLeuGluProIleTyrAspAla 26
 DB 16 GCGCCCGGAAACCTTATATTCAGGCGAGATGACGATACCGAACGAGATATTAATCT 75
 QY 27 TyrValThrAlaLeuSerSerArgMetHisAlaValIle----- 39
 DB 76 ATTCTATATGAGTAATCGCTTTGTAACGCGACGATGAGCGCTCAATAAGTTGAA 135
 QY 40 -----ThrProLysGlyThrSerLeuLysGluSer 49
 DB 136 GGTAGTATTCATTCGAAACATTACATTATGATGAAACGATTAATTGACCGTATATGAGAT 195
 QY 50 ArgIleLysIleSerSerProGlnPheAlaIleGlyGluTyrGluTyrHisIleSerSer 69
 DB 196 AGAATTGAAATCTCAGATGTTCAAGCTGCTTAAGCAA----- 231
 QY 70 AsnThrGluLysProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThrIle 89
 DB 232 -----CTGAATATGTTGTG 246
 QY 90 PheIleValLeuAlaTyrIleGlnProThrGluAlaPheAspLeuGluIleIleTyr 109
 DB 247 ACAAGCTAATGAAAGTGTGAAACAGATGTGCGCAGCTTCATATGAAATTTA----- 297
 QY 110 SerAspProGlyTyrHisSerGlnLysPheThrGluThrLysThrSerSerAsnGlyLys 129
 DB 298 -----AAGCATTTCATTAAACCTTAATGTAACCTTAAGCAGATTAAGCTGCTGCT----- 345
 QY 130 LysThrPheLeuLysThrSerArgAlaIleThrGluValGluLysThrGlyLeuGlySer 149
 DB 346 -----CAGAAGTACGATTAAGTTCA 366
 QY 150 SerAlaGly---LeuValSerValValAlaThrSerLeuLeuSerHisPheIleProAsn 168
 DB 367 AGCGCGCTGTGTTAGTATCTGTGTT-----AAAGCTTGAATGAATCTATAGT----- 417

QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
 DB 418 ---TTGAAATATTCAAACCTTATATTTATTAATAGCTGTAATGCAATATGAAATTA 474
 QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyr---GlyLeuIleVal 207
 DB 475 CAAGTTTAAGTTCATGTTGCT---GATATTCGCGTTAGTACTACAGTGTGCTTGCA 531
 QY 208 TyrArgAspPheGlnProAlaLeuIleAsnPro---ValPheGlnValLeuGluSerAsp 226
 DB 532 TATAGTACGTTTCGAC-----CATACGCGGTGAACAGCAATATGAAAGAA--- 576
 QY 227 ProGluLysPheProThrGluLeuLysLysLeuIleGlnSerAsnTrpGluGluLysHis 246
 DB 577 -----ACATCGGTGAATATGTTTGGAAAAAATTTGCCACGCTTAACAT 621
 QY 247 ---GluArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGly 265
 DB 622 ATCGAACCTTAAACAGCTCCCGAAATATGAAAGTCTTATGATGATGATGCGGCCCA 681
 QY 266 SerGluThrProLysLeuValSerArgValLeuGlnTrpLysGluLysProGluGlu 285
 DB 682 GCTTCTTCTCCACCTTAAGTGAAGTCAACGCTTTAAA----- 723
 QY 286 SerSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGluLeuArg 305
 DB 723 ----- 723
 QY 306 GluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHisSer 325
 DB 724 -----TCAGATCAAGTTTATGATGATGATTTTATGATCAATCT 762
 QY 326 -----ValGluProLeuThrValAlaIleLys-----AsnIle----- 336
 DB 763 CATGCTGTGTGAAGAAAGTTTATATCCAGCTTTTAAACATAATATGAAAGGTGTTCAA 822
 QY 337 -----ArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValPro 350
 DB 823 AAGATGATACGTATTAACAGACGATTTATTTATGATGATTAACGAAATCAAGCTTGA 882
 QY 351 IleGluProAspValGlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysVal 370
 DB 883 ATTGAACAGATTAAGCTTAATAAATTTATGATGCGGTGAAGCAACGTTGGC----- 936
 QY 371 GlyLysValValProGlyAlaGlyIleTyrAsp 381
 DB 937 GCTTCTAATAACTTCAGGTGCTGTGTCGCGAT 969
 RESULT 39
 AEN93055 standard; DNA; 1089 BP.
 AEN93055;
 AC AEN93055;
 XX 24-JUN-2002 (first entry)
 DT XX
 DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2518.
 XX
 DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KW antibacterial; gene therapy; gene; ds.
 XX
 OS Staphylococcus epidermidis.
 XX
 FN US6380370-B1.
 PD 30-APR-2002.
 PD 13-AUG-1998; 98US-0134001.
 PF 14-AUG-1997; 97US-055779B.
 PR 08-NOV-1997; 97US-064964P.

PA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Doucette-Stamm LA, Bush D;
 XX
 XX WPI; 2002-381255/41.
 DR P-PSDB; ABP40510.
 XX
 XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
 PT polypeptide, useful for diagnosing and treating bacterial infections -
 XX
 PS Disclosure; SEQ ID 2518; 267bp; English.
 XX
 XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly S. epidermidis infections. The sequences can be used to
 CC screen for compounds able to interfere with the S. epidermidis life
 CC cycle or inhibit S. epidermidis infection.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.
 XX
 SQ Sequence 1089 BP; 376 A; 159 C; 222 G; 332 T; 0 other;
 Alignment Scores:
 Pred. No.: 6.14e-09 Length: 1089
 Score: 181.50 Matches: 99
 Percent Similarity: 36.74% Conservative: 52
 Best Local Similarity: 24.09% Mismatches: 131
 Query Match: 8.14% Indels: 129
 DB: 24 Gaps: 18
 US-10-069-062-7 (1-432) x ABN93055 (1-1089)
 QY 7 AlaProGlyValLeuAlaLeuValLeuAlaGlyGlyTyrLeuValLeuGluProIleTyrAspAla 26
 DB 28 GCCCGCGGAAACCTTATATATGACGCGAGTATGCAATGCAACGACAGATATTAATCT 87
 QY 27 TyrValThrAlaLeuSerSerArgMetHisAlaValIle----- 39
 DB 88 ATTCTTATTCAGTAATATCGCTTTGTAACGCGACATGAGCGCTCAATAAAGTTGAA 147
 QY 40 -----ThrProGlyGlyThrSerLeuGlySer 49
 DB 148 GGTAGTATTCATCCAAACATTACATTAATGAACCGATTAAATTGACCGTAATGAAGAT 207
 QY 50 ArgIleLeuIleSerSerProGlnPheAlaAsnGlyGluTyrGluTyrHisIleSerSer 69
 DB 208 AGAATTGAATCTCAATGCTTCAAGCTGCTAAGCAA----- 243
 QY 70 AsnThrGluLeuProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThrIle 89
 DB 244 -----CTGAATAATGTTGTG 258
 QY 90 PheIleValLeuAlaTyrIleGlnProThrGluAlaPheAspLeuIleIleIleTyr 109
 DB 259 ACAGCTATGAAGAGTGTGAACAGTATGTCGCGCATGTGCAATATGAAATTA----- 309
 QY 110 SerAspProGlyTyrHisSerGlnGluAspThrGluThrIleThrSerSerAsnGlyGlu 129
 DB 310 -----AAGCAGCTTCACTTAAACATGATGATTAACATTAACATTAACATTAACATTA 357
 QY 130 LysThrPheLeuTyrHisSerArgAlaIleThrGluValGluLeuThrGlyLeuGlySer 149
 DB 358 -----CAGAAATGACGATTAAGTTCA 378
 QY 150 SerAlaGly-----LeuValSerValValAlaThrSerLeuLeuSerHisPheIleProAsn 168
 DB 379 ACGCGCGGTGTGTATATCTGTTGTT-----AAAGCTTTGATTAATTAATTAATTAATTA 429
 QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188

DB 430 ---TTGGAATATATCAAACTTATATATTAATTAATTAATTAATTAATTAATTAATTAATTA 486
 QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyr---GlyLeuIleVal 207
 DB 487 CAAAGTTTAAGTTCAATGTGT---GATATGTGGGTAGTCTACAGTGTGGCTTGCA 543
 QY 208 TyrArgArgPheGlnProAlaLeuIleAsnAsp---ValPheGlnValLeuGluSerAsp 226
 DB 544 TATAGTACGTTGCAC-----CATGCTGGGTGAAACACCAATGAGAA----- 588
 QY 227 ProGluLysPheProThrGluLeuLysLysLeuIleGlnSerAsnThrGluLysHis 246
 DB 589 -----ACATCGGTGATGATGTTTGGAAAAAAATTTGGCCAGGCTTACAT 633
 QY 247 ---GluArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlyGly 265
 DB 634 ATCGAACCTTTACAAAGCTCCGAAATATGAAAGTCTTATTTGATGATGACGCGTCCCA 693
 QY 266 SerGluThrProLysLeuValSerArgValLeuGlnTyrPlyLysGlyLysProGluGlu 285
 DB 694 GCTTCTTCTCCACACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 735
 QY 286 SerSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGluLeuArg 305
 DB 735 ----- 735
 QY 306 GluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHisSer 325
 DB 736 -----TCAGATCCAAAGTTTATTTATGATTTTATGATTTTATGATTTTATGATTTTATGAT 774
 QY 326 -----ValGluProLeuThrValAlaIleLys-----AsnIle----- 336
 DB 775 CATGCTGTGTGAAAGTTTATTCACAGCTTTTAAACTTAATTAATTAATTAATTAATTAATTA 834
 QY 337 -----ArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValPro 350
 DB 835 AAGATGATACGATTAACAGAGATGATTTATTCATCTATGATGATTAACAGATCAGTTGAA 894
 QY 351 IleGluProAspValGlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysVal 370
 DB 895 ATTTGAACAGATTAACCTTAATAAATATATGATGTGCGGTGAAGACACGCTGGC----- 948
 QY 371 GlyGlyValValProGlyValGlyGlyTyrAsp 381
 DB 949 GCTTCAAACTTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 981
 RESULT 40
 AAHS4932/c
 ID AAHS4932 standard; DNA; 4010 BP.
 AC AAHS4932;
 XX
 XX 03-SEP-2001 (first entry)
 DT XX
 XX
 DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:4296.
 XX
 XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis; ds.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN WO200134809-A2.
 XX
 XX 17-MAY-2001.
 PD XX
 PD 09-NOV-2000; 2000WO-US30782.
 PF XX
 PR 09-NOV-1999; 99US-0164258.
 XX
 PA (GLAXO) GLAXO GROUP LTD.
 XX
 XX Kimmerly WJ;

XX WPI; 2001-316495/33.
 DR Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis -
 XX
 XX Claim 8; Page 2029-2030; 2188pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AG81454 to AG81120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce host cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH5091 to
 CC AAH5098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 4010 BP; 1230 A; 806 C; 556 G; 1418 T; 0 other;

Alignment Scores:
 Pred. No.: 3, 91e-08 Length: 4010
 Score: 181.50 Matches: 99
 Percent Similarity: 36.74% Conservative: 52
 Best Local Similarity: 24.09% Mismatches: 131
 Query Match: 8.14% Indels: 129
 DB: 22 Gaps: 18

US-10-069-062-7 (1-432) x AAH54932 (1-4010)

QY 7 AlaProGlyLysAlaHeuLeuAgiGlyTyrLeuValLeuGluProIleTyrAspAla 26
 DB 1156 GCGCCCGGAAACCTTATATGTCAGGCGAGTATGACGTAACCGAATGATATTAATCT 1097
 QY 27 TyrValThrAlaLeuSerSerArgMetHisAlaValIle----- 39
 DB 1096 ATCTTATTCAGTAATATGCTTTGTATACGCGACAAATGAGCGCTCAATTAAGTTGAA 1037
 QY 40 -----ThrProLeuGlyThrSerLeuLysGluSer 49
 DB 1036 GGTAGTATTCATCCAAACATTACATTATGAAACCGATTAAATTGACCGTATATGAAGT 977
 QY 50 ArgIleLysIleSerSerProGlnPheAlaAsnGlyGluTrpGluTyrHisIleSerSer 69
 DB 976 AGAATTGAATCTCAGATGTCAGCTGTAAGCAA----- 941
 QY 70 AsnThrGluLysProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThrIle 89
 DB 940 -----CTGAATATGTGTG 926
 QY 90 PheIleValLeuAlaTyrIleGlnProThrGluAlaPheAspLeuGluIleIleTyr 109
 DB 925 ACAGCATATGAACTGTTTAAACAGTATGCGCAGTTCGCAATATGAAATTA----- 875
 QY 110 SerAspProGlyTyrHisSerGlnLeuAspThrGluTyrThrLysSerSerAsnGlyGlu 129
 DB 874 -----AAGCATTTCAATTAACCATTAAGTACTTACGATTAAGTCTGCT----- 827
 QY 130 LysThrPheLeuTyrHisSerArgAlaIleThrGluValGlnLysThrGlyLeuGlySer 149
 DB 826 -----CAGAGTACGATTAAGTTCA 806

QY 150 SerAlaGly---LeuValSerValValAlaThrSerLeuLeuSerHisPheIleProAsn 168
 DB 805 AGCGCGCGCTTTTATGATCTCTGTT-----AAAGCTTTGAATGAATTCATAGGT----- 755
 QY 169 ValIleSerThrAsnLysPheAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
 DB 754 ---TTGAAATATATCAAAACCTTATATTAATATGATGTAATGCAAAATATGAATTA 698
 QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyr---GlyLeuIleVal 207
 DB 697 CAAAGTTTAAGTTCATGTCGCG--GATATGCGGTTTATGATGTCATGTCGTCGCTTGA 641
 QY 208 TyrArgAspPheGlnProAlaLeuIleAsnAsp---ValPheGlnValLeuGluSerAsp 226
 DB 640 TATAGTACGTTTGAC-----CATGCTGCGTGGAACACCAATATGAAGAA----- 596
 QY 227 ProGluLysPheProThrGluLeuLysLysLeuIleGluSerAsnTyrGluLysHis 246
 DB 595 -----ACATCGGTGAATGATGTTTGGAAAAAATTTGGCCAGGCTTACAT 551
 QY 247 ---GluArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlyGly 265
 DB 550 ATCGAACCTTTTCAAGCTCCGAAATATATGAAAGTCCTTATGATGATGACCTGGTCTCCA 491
 QY 266 SerGluThrProLysLeuValSerArgValLeuGlnIleTyrLysGluLysProGluGlu 285
 DB 490 GCTTCTTCCACACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 449
 QY 286 SerSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGluLeuArg 305
 DB 449 ----- 449
 QY 306 GluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHisSer 325
 DB 448 -----TCAGATCCAGCTTTTATGATGATGATGATGATGATGATGATGATGATGAT 410
 QY 326 -----ValGluProLeuThrValAlaIleLys-----AsnIle----- 336
 DB 409 CATGCTGTGTGTAAGTATTAATCCAGCTTTTAAATTAATTAATTAATTAATTAATTAATTA 350
 QY 337 -----ArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValPro 350
 DB 349 AAGATATATCATATTAACAGACGATATTAATCAATCAATCAATCAATCAATCAATCAATCA 290
 QY 351 IleGluProAspValGlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysVal 370
 DB 289 ATTGAAACAGATTAAGCTTAAAAAATTAATGATGATGATGATGATGATGATGATGATGAT 236
 QY 371 GlyGlyValValProGlyAlaGlyGlyTyrAsp 381
 DB 235 GCTTCAAAACCTTCAGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 203

Search completed: August 15, 2003, 14:16:43
 Job time : 414 sec

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: August 8, 2003, 09:47:37 ; Search time 99 seconds
(without alignments)
1126.048 Million cell updates/sec

Title: US-10-069-062-7
Perfect score: 432
Sequence: 1 MSKAPSAKGAFLAGYLVL.....DLEQTEGVLEKPEYDYL 432

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvins:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	2.1	2463	12 Q8BES6	Q8BES6 bovine kobu
2	8	1.9	173	10 Q40650	Q40650 Oryza sativ
3	8	1.9	312	5 Q93482	Q93482 caenorhabdi
4	8	1.9	340	16 Q8XDH9	Q8XDH9 escherichia
5	8	1.9	340	16 Q8FII3	Q8FII3 escherichia
6	8	1.9	374	10 Q9ST01	Q9ST01 arabidopsis
7	8	1.9	419	10 Q8H1R7	Q8H1R7 arabidopsis
8	8	1.9	426	3 Q9UT88	Q9UT88 schizosach
9	8	1.9	465	10 Q8VYQ8	Q8VYQ8 arabidopsis
10	8	1.9	484	10 Q9C8P9	Q9C8P9 arabidopsis
11	8	1.9	496	17 Q9TVL8	Q9TVL8 sulfolobus
12	8	1.9	503	10 Q944G1	Q944G1 hevea bras
13	8	1.9	505	10 Q9C6T1	Q9C6T1 arabidopsis
14	8	1.9	505	16 Q9K7C5	Q9K7C5 bacillus ha
15	8	1.9	510	4 Q9H4V1	Q9H4V1 homo sapien
16	8	1.9	568	2 Q44106	Q44106 anaplasma m

17	8	1.9	645	16 Q9PNH8	Q9PNH8 campylobact
18	8	1.9	770	5 Q8WQ41	Q8WQ41 leishmania
19	8	1.9	780	5 Q9GYD6	Q9GYD6 leishmania
20	8	1.9	953	16 P73173	P73173 synecocyst
21	8	1.9	1028	4 Q60283	Q60283 homo sapien
22	8	1.9	1161	10 Q9LR20	Q9LR20 arabidopsis
23	8	1.9	3396	5 Q62263	Q62263 caenorhabdi
24	8	1.9	27	2 Q54783	Q54783 streptococc
25	7	1.6	72	16 Q8XIC9	Q8XIC9 clostridium
26	7	1.6	73	16 Q8R5T5	Q8R5T5 thermococ
27	7	1.6	75	16 Q928F1	Q928F1 listeria in
28	7	1.6	80	12 Q89456	Q89456 hepatitis g
29	7	1.6	94	16 Q9PB81	Q9PB81 xyella fas
30	7	1.6	99	12 Q92437	Q92437 bombyx mori
31	7	1.6	99	12 Q8B9H9	Q8B9H9 rachiplusia
32	7	1.6	102	10 P82726	P82726 arabidopsis
33	7	1.6	103	15 Q98WV2	Q98WV2 human immun
34	7	1.6	110	5 Q9U1X9	Q9U1X9 caenorhabdi
35	7	1.6	113	15 Q9E487	Q9E487 human immun
36	7	1.6	113	15 Q9E488	Q9E488 human immun
37	7	1.6	113	15 Q9E491	Q9E491 human immun
38	7	1.6	116	8 Q9B8A4	Q9B8A4 trichinella
39	7	1.6	124	10 Q8LCV8	Q8LCV8 arabidopsis
40	7	1.6	129	16 Q84225	Q84225 chlamydia t
41	7	1.6	137	17 Q9YE18	Q9YE18 aeropyrum p
42	7	1.6	151	16 Q8ED72	Q8ED72 shewanella
43	7	1.6	158	11 Q8CF18	Q8CF18 mus musculu
44	7	1.6	160	2 Q05683	Q05683 mycobacteri
45	7	1.6	168	16 Q9KN58	Q9KN58 vibrio chol

ALIGNMENTS

RESULT 1

ID	Q8BES6	PRELIMINARY;	PRT; 2463 AA.
AC	Q8BES6;		
DT	01-MAR-2003 (TREMBLrel. 23, Created)		
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)		
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)		
DE	Genomic RNA, complete genome, strain:U-1.		
OS	Bovine kobuvirus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;		
OC	Kobuvirus.		
OX	NCBI_TaxID=194965;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=U-1;		
RA	Yamashita T., Ito M., Kabashima Y., Tsuzuki H., Sakae K.;		
RT	"ISOLATION AND CHARACTERIZATION OF A NEW SPECIES OF KOBUVIRUS		
RT	ASSOCIATED WITH BOVINE."		
RL	Submitted (May-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB084788; BAC23066.1; -		
FT	CHAIN	1	187
FT	CHAIN	188	554
FT	CHAIN	555	777
FT	CHAIN	778	1044
FT	CHAIN	1045	1178
FT	CHAIN	1179	1343
FT	CHAIN	1344	1678
FT	CHAIN	1679	1772
FT	CHAIN	1773	1802
FT	CHAIN	1803	1994
FT	CHAIN	1995	2463
SO	SEQUENCE	2463 AA;	270562 MW; AB2CF6574BCCA2 CRC64;

Query Match 2.1%; Score 9; DB 12; Length 2463;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

398 KTLKPDYF 406
|||||||

Db 2140 KTLNPDYF 2148

RESULT 2

Q040650 PRELIMINARY; PRT; 173 AA.
 ID Q040650
 AC Q040650
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Wilm's tumor-related protein QM (Fragment).
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriarthroideae; Oryzaceae; Oryza.
 OC NCBI_TaxID=4530;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Indica variety Guang-Lu-Ai No.4; TISSUE=Seedling;
 RA Zong H., Jiang Y., Cao K.;
 RT "Isolation and Characterization of a rice QM gene, tumor suppressor or
 development regulator";
 RL Thesis (1996), Biochemistry, Fudan University.
 DR EMBL; U55212; AAA99158.1; -.
 DR Gramene; Q040650; -.
 DR InterPro; IPR001197; Ribosomal L10E.
 DR Pfam; PF00826; Ribosomal L10E; 1.
 DR PROSITE; PS01257; RIBOSOML_L10E; 1.
 FT NON TER 1 1
 SQ SEQUENCE 173 AA; 19262 MW; CEA0448DCAE5DB9 CRC64;

Query Match 1.9%; Score 8; DB 10; Length 173;
 Best Local Similarity 100.0%; Pred.No.12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 APGKAFIA 14
 159 APGKAFIA 166

RESULT 3
 Q03482 PRELIMINARY; PRT; 312 AA.
 ID Q03482
 AC Q03482
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 GN F14H8.2 protein.
 GN F14H8.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematozoa; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Matthews L.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 investigating biology";
 RL Science 282:2012-2018(1998).
 DR EMBL; Z81061; CAB02931.1; -.
 DR Wormpep; F14H8.2; CE09394.
 SQ SEQUENCE 312 AA; 36148 MW; CB98CBF96BCB65D3 CRC64;

Query Match 1.9%; Score 8; DB 5; Length 312;
 Best Local Similarity 100.0%; Pred.No.21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 233 ELKKLIES 240

Db 73 ELKKLIES 80

RESULT 4

Q08XD9 PRELIMINARY; PRT; 340 AA.
 ID Q08XD9
 AC Q08XD9
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE 2166 protein (Putative ATP-binding protein of ABC transporter).
 GN 2166 OR ECS1699.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OC NCBI_TaxID=83334;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 Grothbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 Apodaca J., Anantharaman T.S., Lin J., Yen J., Schwartz D.C.,
 Welch R.A., Blattner F.R.;
 RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533(2001).
 RT [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 Han C.-g., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tode T.,
 Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RA "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:111-22(2001).
 DR EMBL; AE005337; AAG56054.1; -.
 DR EMBL; AP002556; BAB35122.1; -.
 DR InterPro; IPR002491; Peripla BP.
 DR Pfam; PF01497; Peripla_BP_2; 1.
 KM Complete proteome.
 SQ SEQUENCE 340 AA; 37641 MW; FFE547B8CFD556B CRC64;

Query Match 1.9%; Score 8; DB 16; Length 340;
 Best Local Similarity 100.0%; Pred.No.22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 LHNVAQIA 184
 219 LHNVAQIA 226

RESULT 5

Q08FI13 PRELIMINARY; PRT; 340 AA.
 ID Q08FI13
 AC Q08FI13
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Conserved hypothetical protein.
 GN C1652.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OC NCBI_TaxID=217992;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasbo D., Buckles E.L., Liu S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic *Escherichia coli*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 DR EMBL; AE016759; AAN80117.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 340 AA; 37690 MW; 032B627796698F27 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 340;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 LHMVAQIA 184
 DB 219 LHMVAQIA 226

RESULT 6

Q9ST01 PRELIMINARY; PRT; 374 AA.

AC Q9ST01; 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Hypothetical 42.3 kDa protein.
 GN T25K17.150 OR AT4G26340.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsie.
 NCBI_TaxID=3702;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Koester P., Hempel S., Entian K.-D., Baneroff I.,
 RA Mewes H.W., Mayer K.F.X., Schueller C.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Rose M., Hempel S., Entian K.-D., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL049171; CAB38963.1; -.
 DR EMBL; AL161565; CAB79489.1; -.
 DR InterPro; IPR001810; F-box.
 DR InterPro; IPR006566; F-box.
 DR Pfam; PF00646; F-box; 1.
 DR SMART; SM00579; F-box; 1.
 DR SMART; SM00256; F-box; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 374 AA; 42262 MW; 391638BFD39670D CRC64;

Query Match 1.9%; Score 8; DB 10; Length 374;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 VVATSLLS 163
 DB 24 VVATSLLS 31

RESULT 7
 Q8H1R7 PRELIMINARY; PRT; 419 AA.

AC Q8H1R7;
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN AT4G26340.

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsie.
 NCBI_TaxID=3702;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,
 RA Lee J.M., Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C.,
 RA Wu H.C., Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J.,
 RA Nguyen M., Paim C.J., Shinn P., Southwick A., Tripp M.G., Wu T.,
 RA Davis R.W., Becker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones."
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY142504; AAN13055.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 419 AA; 47300 MW; 0B44E17262146B3 CRC64;

Query Match 1.9%; Score 8; DB 10; Length 419;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 VVATSLLS 163
 DB 24 VVATSLLS 31

RESULT 8

Q9UT88 PRELIMINARY; PRT; 426 AA.

AC Q9UT88;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Putative phosphomethyltransferase kinase.
 GN SPAC343.01C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 NCBI_TaxID=4896;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=97zh-;
 RA Murphy L., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL109739; CAB52264.1; -.
 DR GeneDB; SPombe; SPAC343.01C; -.
 DR InterPro; IPR005916; Pmev_kin_erg8.
 DR TIGRFAMs; TIGR01219; Pmev_kin_erg8; 1.
 KW Kinase.
 SQ SEQUENCE 426 AA; 47322 MW; 25635B99E7FPAAD64 CRC64;

Query Match 1.9%; Score 8; DB 3; Length 426;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 KTGAGSSA 151
 DB 153 KTGAGSSA 160

RESULT 9
 Q8VY08 PRELIMINARY; PRT; 465 AA.
 AC Q8VY08;
 DT 01-MAR-2002 (TREMblrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN AT1G35470.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC euroside II; Brassicales; Brassicaceae; Arabidopsis.
 NC NCB1_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banb J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,
 RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carinici P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shinn P., Southwick A., Shimozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Full length cDNA of gene At1g35470 (GI:15219607).";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banb J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carinici P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shimozaki K.,
 RA Davis R.W., Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY070090; AA149784.1; -;
 DR EMBL; AY06710; AA020344.1; -;
 DR InterPro; IPR003877; SPRY_receptor.
 DR Pfam; PF00622; SPRY; 1.
 KM Hypothetical protein..
 SQ SEQUENCE 465 AA; 52093 MW; 988BA20DE9F31156 CRC64;
 QY
 DB 281 EXPRESSV 288
 383 EXPRESSV 390
 PRT; 484 AA.
 RESULT 10
 Q9C8P9 PRELIMINARY;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical 54.0 kDa protein.
 GN F12A.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC euroside II; Brassicales; Brassicaceae; Arabidopsis.
 NC NCB1_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Elugu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,

RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenskaia I., Kutz D.B., Kwan A., Lam B.,
 RA Langin-Hopfer S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maier R., Maritali A.,
 RA Millischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vayberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana";
 RL Nature 408:816-820(2000).
 DR EMBL; AC023064; AAG52111.1; -;
 DR InterPro; IPR006595; CTLH.
 DR InterPro; IPR006594; LISH.
 DR InterPro; IPR003877; SPRY_receptor.
 DR Pfam; PF00622; SPRY; 1.
 DR SMART; SM00668; CTLH; 1.
 DR SMART; SM00667; LISH; 1.
 KM Hypothetical protein..
 SQ SEQUENCE 484 AA; 54010 MW; 65AC5A0D527766FC CRC64;
 QY
 DB 281 EXPRESSV 288
 402 EXPRESSV 409
 PRT; 496 AA.
 RESULT 11
 Q97VL8 PRELIMINARY;
 ID Q97VL8
 AC Q97VL8;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE ABC transporter, ABC1 family, putative.
 GN SSO2605.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 NC NCB1_Taxid=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; PubMed=11427726;
 RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtiss B.A.,
 RA De Moers A., Erasuo G., Fletcher C., Gordon P.M.K.,
 RA Heikamp-de Jong I., Jeffries A.C., Kozeira C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 DR EMBL; AE006856; AA042726.1; -;
 DR InterPro; IPR004147; ABC1.
 DR InterPro; IPR000479; Prot_kinase.
 DR Pfam; PF03109; ABC1; 1.
 DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
 KM Complete proteome..
 SQ SEQUENCE 496 AA; 57515 MW; DF5170C21E3DCEBB CRC64;
 QY
 DB 315 PFTYIKEL 322
 70 PFTYIKEL 77
 PRT; 322 AA.
 Query Match 1.9%; Score 8; DB 17; Length 496;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 12
ID 0944G1 PRELIMINARY; PRT; 503 AA.
AC 0944G1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Phosphomevalonate kinase.
OS Hevea brasiliensis (Para rubber tree).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Euphorbiaceae; Hevea.
OX NCBI_TaxID=3981;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. RRM600;
RA Hallahan D.L., Keiper-Hrynko N.M.;
RT "Genes involved in the biosynthesis of isopentenyl diphosphate in the
RT rubber tree Hevea brasiliensis.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AA429385; AAL18926.1; -.
DR InterPro; IPR005916; Pmey_kin_erg8.
DR TIGRFAMs; TIGR01219; Pmey_kin_erg8; 1.
KW Kinase.
SQ SEQUENCE 503 AA; 54171 MW; BE6F91B0845FF94 CRC64;

Query Match 1.9%; Score 8; DB 10; Length 503;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 KTGLGSSA 151
DB 180 KTGLGSSA 187

RESULT 13
ID 09C6T1 PRELIMINARY; PRT; 505 AA.
AC 09C6T1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Hypothetical 54.4 kDa protein.
GN F5M6.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chun A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Eguu P., Feldblyum T.V., Feng J.-D., Feng B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetska I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Malt R., Marshall A.,
RA Millscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz S.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vayenberg M., Vysotskaya V.S., Walker M.,
RA Wu D., Yu G., Frazer C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.";

```

```

RL Nature 408:816-820(2000).
DR EMBL; AC079041; AAG50716.1; -.
DR InterPro; IPR005916; Pmey_kin_erg8.
DR TIGRFAMs; TIGR01219; Pmey_kin_erg8; 1.
KW Hypothetical protein.
SQ SEQUENCE 505 AA; 54409 MW; B1CBAC6A338B3D63 CRC64;

Query Match 1.9%; Score 8; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 KTGLGSSA 151
DB 179 KTGLGSSA 186

RESULT 14
ID 09K7C5 PRELIMINARY; PRT; 505 AA.
AC 09K7C5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Sodium-dependent transporter.
GN BH3439.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kohara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001518; BAB07158.1; -.
DR InterPro; IPR000175; Na/ntan_sympor.
DR Pfam; PF00209; SNF; 1.
DR PRINTS; PR00176; NANEUSMPORT.
DR ProDom; PD000448; Na/ntan_sympor; 1.
DR PROSITE; PS00610; NA_NEUOTRAN_SYMP_1; 1.
DR PROSITE; PS50267; NA_NEUOTRAN_SYMP_3; 1.
KW Complete proteome.
SQ SEQUENCE 505 AA; 54296 MW; 607C0F6A00D67B07 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 505;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 AGLVSVVA 158
DB 393 AGLVSVVA 400

RESULT 15
ID 09H4V1 PRELIMINARY; PRT; 510 AA.
AC 09H4V1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE BA33811.1 (Novel CUB domain protein similar to attractin)
DE (Fragment).
GN BA33811.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

RA Bird C.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL; AL355530; CAC12966.1; -.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006652; Kelch_rep.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF01344; Kelch; 2.
DR SMART; SM00042; CUB; 1.
DR SMART; SM00181; EGF; 2.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 1.
DR EGF-like domain.
FT NON_TER 510 510
SQ SEQUENCE 510 AA; 56370 MW; 1C45C837DCB2D7E9 CRC64;

Query Match 1.9%; Score 8; DB 4; Length 510;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 EYHISNT 71
DB 446 EYHISNT 453

RESULT 16
ID Q4106 PRELIMINARY; PRT; 568 AA.
AC Q4106;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE Major surface protein 3-19 (Fragment).
GN MSP3-19.
OS Anaplasma marginale.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Anaplasma.
OX NCBI_TaxID=770;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Florida;
RX MEDLINE=97130030; PubMed=8975906;
RA Allen A.R., Palmer G.H., McGuire T.C., McElwain T.F., Perryman L.E.,
RA Barbet A.F.,
RT "Anaplasma marginale major surface protein 3 is encoded by a
RT polymorphic, multigene family."
RL Infect. Immun. 65:156-163(1997).
DR EMBL; U60779; AAB49276.1; -.
DR HSP; P03069; 1YSA.
FT NON_TER 1 1
SQ SEQUENCE 568 AA; 63463 MW; AFP3E9F1B4DF66A3 CRC64;

Query Match 1.9%; Score 8; DB 2; Length 568;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 ELREMERK 310
DB 169 ELREMERK 176

RESULT 17
ID Q9PNH8 PRELIMINARY; PRT; 645 AA.
AC Q9PNH8;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE Membrane bound zinc metalloproteinase (EC 3.4.24.-).
GN FTSH OR Ctl116C.

OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Tagatz K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.M.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrall B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
DR EMBL; AL139077; CAB73371.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR InterPro; IPR003960; AAA_sub.
DR InterPro; IPR006642; Peptidase_M41.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF01344; Peptidase_M41; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR01241; FtsH_fam; 1.
DR PROSITE; PS00674; AAA; 1.
KW Complete proteome.
SQ SEQUENCE 645 AA; 71048 MW; 0A68B664887BC02 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 645;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ELKLIIES 240
DB 64 ELKLIIES 71

RESULT 18
ID Q8WQ41 PRELIMINARY; PRT; 770 AA.
AC Q8WQ41;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical predicted WD40 repeat protein P265.09, unknown
DE function.
GN P265.09.
OS Leishmania major.
OC Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RA Robben J., Grymompres B., Weltjens I., Aert R., Voicaert G.,
RA Ivens A.C., Quail M., Rajandream M.A., Barrall B.G.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome."
RL Genome Res. 8:135-145(1998).
DR EMBL; AL359716; CAD19418.1; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 3.
DR SMART; SM00320; WD40; 2.
KW Repeat, WD repeat.
SQ SEQUENCE 770 AA; 82530 MW; 4E9AF5A321AF047 CRC64;

Query Match 1.9%; Score 8; DB 5; Length 770;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 GSSAGLVS 155
 |||||
 Db 461 GSSAGLVS 468

RESULT 19

Q9GYD6 PRELIMINARY; PRT; 780 AA.

AC Q9GYD6; 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical 83.9 kDa protein.
 GN L3180.01.
 OS Leishmania major.
 OC Eukaryota; Euzoenzoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Friellin;
 RA Oliver K., Murphy L., Harris D., Ivens A.C., Quail M.,
 RA Rajandream M.A., Barrell B.G.;
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Friellin;
 RX MEDLINE=98146435; PubMed=9477341;
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
 RA Smith D.F.;
 RT "A physical map of the Leishmania major Friellin genome."
 RL Genome Res. 8:135-145(1998).
 DR EMBL; AL391629; CAC0307.1; -.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 3.
 KW Hypothetical protein; Repeat; WD repeat.
 SQ SEQUENCE 780 AA; 83875 MW; 254D60BCAD1C6E CRC64;

Query Match 1.9%; Score 8; DB 5; Length 780;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 GSSAGLVS 155
 |||||
 Db 471 GSSAGLVS 478

RESULT 20

P73173 PRELIMINARY; PRT; 953 AA.

AC P73173; 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Methyl-accepting chemotaxis protein (MCP).
 GN PILD OR SLI1294.
 OS Synecocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=97061201; PubMed=8905231;
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirosewa M., Sugita M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuo A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synecocystis sp. strain PCC6803. II. Sequence determination of the
 entire genome and assignment of potential protein-coding regions."

RL DNA Res. 3:109-136(1996).
 DR EMBL; D90904; BA017199.1; -.
 DR HSSP; P02942; 1QU7.
 DR InterPro; IPR004089; Chmtaxis_transd.
 DR InterPro; IPR003660; HAMF.
 DR InterPro; IPR004090; Me_chemotaxis.
 DR Pfam; PF00672; HAMF; 1.
 DR Pfam; PF00015; MCPsignal; 1.
 DR PRINTS; PR00260; CHEMTRNSDUCR.
 DR SMART; SM00304; HAMF; 2.
 DR SMART; SM00283; MA; 1.
 DR PROSITE; PS01111; CHEMOTAXIS_TRANSDUC_2; 1.
 KW Complete proteome.

SQ SEQUENCE 953 AA; 103178 MW; A274B02DBEF9CBB CRC64;

Query Match 1.9%; Score 8; DB 16; Length 953;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 EPLTVAIK 334
 |||||
 Db 491 EPLTVAIK 498

RESULT 21

ID 060283 PRELIMINARY; PRT; 1028 AA.

AC 060283; 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE KIA0534 protein (Fragment).
 GN KIA0534.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98290545; PubMed=9628581;
 RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
 RA Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 The complete sequences of 100 new cDNA clones from brain which can
 code for large proteins in vitro."
 RT DNA Res. 5:31-39(1998).
 RL EMBL; AB011106; BA025460.2; -.
 DR HSSP; P02468; 1KLO.
 DR InterPro; IPR006652; Kelch_rep.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001304; Lactin_C.
 DR InterPro; IPR003659; Plexin-like.
 DR InterPro; IPR002165; Plexin_repeat.
 DR Pfam; PF01344; Kelch; 4.
 DR Pfam; PF00053; laminin_EGF; 1.
 DR Pfam; PF00059; lectin_C; 1.
 DR Pfam; PF01437; PSI; 4.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00180; EGF_Lam; 2.
 DR SMART; SM00423; PSI; 5.
 DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
 KW Laminin EGF-like domain.
 FT NON_TER
 SQ SEQUENCE 1028 AA; 114203 MW; 45504CF973700D06 CRC64;

Query Match 1.9%; Score 8; DB 4; Length 1028;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 EYHISNT 71
 |||||

Db 95 EYHISNT 102

RESULT 22

ID Q9LRZ0 PRELIMINARY; PRT; 1161 AA.

AC Q9LRZ0; 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Genomic DNA, chromosome 3, TNC clone:K2019.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucosids II; Brassicales; Brassicaceae; Arabidopsis

NCBI_TaxID=3702;

[1]

SEQUENCE FROM N.A.

RC STRAIN=Columbia;

RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;

RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

[2]

SEQUENCE FROM N.A.

RC STRAIN=Columbia;

RX MEDLINE=20277480; PubMed=10819329;

RA Nakamura Y.;

RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence

RT features of the regions of 4,504,864 bp covered by sixty P1 and TNC

RT clones.";

RL DNA Res. 7:131-135(2000).

DR EMBL; AB028608; BAA95777.1; -.

DR InterPro; IPR006595; CTIH.

DR InterPro; IPR006594; LISH.

DR InterPro; IPR001680; WD40.

DR Pfam; PF00400; WD40; 10.

DR SMART; SM00668; CTIH; 1.

DR SMART; SM00667; LISH; 1.

DR SMART; SM00320; WD40; 8.

DR PROSITE; PS00678; WD REPEATS 1; 1.

DR PROSITE; PS50082; WD REPEATS 2; 2.

DR PROSITE; PS50294; WD_REPEATS_REGION; 2.

KW Repeat; WD repeat.

SQ SEQUENCE 1161 AA; 128027 MW; 4F2BC3AEBB0A55A0 CRC64;

Query Match

Best Local Similarity 1.9%; Score 8; DB 10; Length 1161;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 TELKLIIE 239

Db 145 TELKLIIE 152

RESULT 23

ID Q62263 PRELIMINARY; PRT; 3396 AA.

AC Q62263; 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE F54B3.1 protein.

GN F54B3.1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

NCBI_TaxID=6239;

[1]

SEQUENCE FROM N.A.

RP Murrin A. A.;

RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

[2]

SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for

RT investigating biology.";

RL Science 282:2012-2016(1998).

DR EMBL; Z48583; CAA8472.1; -.

DR WormPep; F54B3.1; CE17865.

SQ SEQUENCE 3396 AA; 391875 MW; 337F20A3A9BD2CE7 CRC64;

Query Match

Best Local Similarity 1.9%; Score 8; DB 5; Length 3396;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 QIAHCTAQ 189

Db 1216 QIAHCTAQ 1223

RESULT 24

ID Q54783 PRELIMINARY; PRT; 27 AA.

AC Q54783; 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)

DE Pneumococcal transposon Tn1545 aphA-3 gene n (Fragment).

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

NCBI_TaxID=1313;

[1]

SEQUENCE FROM N.A.

RX MEDLINE=87286391; PubMed=3039302;

RA Galland F., Tien-Cuot P., Carlier C., Courvalin P.;

RT "Nucleotide sequence of the kanamycin resistance determinant of the

RT pneumococcal transposon Tn1545: Evolutionary relationships and

RT transcriptional analysis of aphA-3 genes.";

RL Mol. Gen. Genet. 207:509-513(1987).

DR EMBL; X05577; CAA29083.1; -.

FR NON TER

SQ SEQUENCE 27 AA; 3186 MW; 6017E09ADC9EF455 CRC64;

Query Match

Best Local Similarity 1.6%; Score 7; DB 2; Length 27;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ELKLIIE 239

Db 9 ELKLIIE 15

RESULT 25

ID Q8XIC9 PRELIMINARY; PRT; 72 AA.

AC Q8XIC9; 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE ATP synthase C chain.

GN ATP OR CPE2192.

OS Clostridium perfringens.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

NCBI_TaxID=1502;

[1]

SEQUENCE FROM N.A.

RC STRAIN=13 / Type A;

RX PubMed=11792842;

RA Shimizu T., Ohtani K., Hirakawa H., Oshima K., Yamashita A.;

RA Shiba T., Ogasawara N., Hatton M., Kuhara S., Hayashi H.;

RT "Complete genome sequence of Clostridium perfringens, an anaerobic

RT flesh-eater.";

Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).

DR EMBL; AP003193; BAB81898.1; -.

DR InterPro; IPR002379; ATPase_Csub.
 DR InterPro; IPR005953; ATP_synth_C.
 DR InterPro; IPR000454; Sub ATPase_Csub.
 DR Pfam; PF00137; ATP_synth_C; 1.
 DR PRINTS; PR00124; ATPaseC.
 DR TIGRPFAM; TIGR01260; ATP_synth_C; 1.
 DR PROSITE; PS00605; ATPase_C; 1.
 KW Complete proteome.
 SQ SEQUENCE 72 AA; 7087 MW; DA813MA2P5A5B6A8 CRC64;

Query Match
 Best Local Similarity 1.6%; Score 7; DB 16; Length 72;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 ATAIYGL 205
 |||||
 DB 56 ATAIYGL 62

RESULT 26
 O8R5T5 PRELIMINARY; PRT; 73 AA.

AC O8R5T5;
 DT 01-JUN-2002 (TRENBLrel. 21, Created)
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE FOPI-type ATP synthase c subunit/archaeal/vacuolar-type H-ATPase
 DE subunit K.
 GN ATP OR TTE0631.
 OS Thermoanaerobacter tengcongensis.
 OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
 OC Thermoanaerobacteriaceae; Thermoanaerobacter.
 OX NCBI_TaxId=119072;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=MB4 / JCM 11007;
 RX MEDLINE=21992816; PubMed=11997336;
 RA Bao Q., Tian Y., Li W., Xu Z., Hu S., Dong W., Yang J.,
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
 RA Tan H., Chen R., Wang J., Yu J., Yang H.,
 RT "A complete sequence of T. tengcongensis genome."
 RL Genome Res. 12:689-700(2002).
 DR EMBL; AE013032; AAM23900.1; -.
 DR InterPro; IPR002379; ATPase_Csub.
 DR InterPro; IPR005953; ATP_synth_C.
 DR InterPro; IPR000454; Sub ATPase_Csub.
 DR Pfam; PF00137; ATP_synth_C; 1.
 DR TIGRPFAM; TIGR01260; ATP_synth_C; 1.
 DR PROSITE; PS00605; ATPase_C; 1.
 KW Complete proteome.
 SQ SEQUENCE 73 AA; 7159 MW; 308AC096F8CB99 CRC64;

Query Match
 Best Local Similarity 1.6%; Score 7; DB 16; Length 73;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 ATAIYGL 205
 |||||
 DB 56 ATAIYGL 62

RESULT 27

O928F1 PRELIMINARY; PRT; 75 AA.

AC O928F1;
 DT 01-DEC-2001 (TRENBLrel. 19, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
 DE Hypothetical protein lin2584.
 GN LIN2584.
 OS Listeria innocua.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 OX NCBI_TaxId=1642;

RM [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIP 11262 / Serovar 6a;
 RX PubMed=11679669;
 RA Glaser P., Frangeul L., Buchrieser C., Rumniock C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetoui F., Couve E., de Daruvar A., Deloux P.,
 RA Dommann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,
 RA Entian K.-D., Fathi H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kuntz F., Kurapkak G.,
 RA Madieno E., Maitournam A., Mata Vicente J., Ng E., Nedjati H.,
 RA Nordisk G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schueter T., Simoes N., Tietz A.,
 RA Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.,
 RT "Comparative genomics of Listeria species."
 RL Science 294:849-852(2001).
 DR EMBL; AL596172; CAC97811.1; -.
 DR ListList; LIN02584; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 75 AA; 9248 MW; 0ED543ADEB95A81 CRC64;

Query Match
 Best Local Similarity 1.6%; Score 7; DB 16; Length 75;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ELKLLIE 239
 |||||
 DB 57 ELKLLIE 63

RESULT 28

O89456 PRELIMINARY; PRT; 80 AA.

AC O89456;
 DT 01-NOV-1998 (TRENBLrel. 08, Created)
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Helicase (Fragment).
 OS Hepatitis GB virus C.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC GBV-C/HGV group.
 OX NCBI_TaxId=39839;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GBV-C/S-lym40;
 RA Seipp S., Wahl R., Goesser T., Theilmann L., Stremmel W.,
 RA Kallinowski B.,
 RT "Prevalence of GB virus C (GBV-C) in Human Peripheral Blood
 RT Mononuclear Cells (PBMC)."
 RT Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF019211; AAC31796.1; -.
 DR MEROPS; S29.002; -.
 DR MEROPS; U39.001; -.
 FT NON TER 1 1
 FT NON TER 80 80
 SQ SEQUENCE 80 AA; 8572 MW; 7994AB1E784394E CRC64;

Query Match
 Best Local Similarity 1.6%; Score 7; DB 12; Length 80;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TALSSRM 35
 |||||
 DB 56 TALSSRM 62

RESULT 29

O9PBS1 PRELIMINARY; PRT; 94 AA.

AC O9PBS1;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)

DT 01-MAR-2002 (T-EMBLrel. 20, last annotation update)
 DE Hypothetical protein xf2069.
 GN XF2069.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 ON NCB1_TaxID=2371;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9a5c;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvares A.R., Alves L.M.C., Araya J.E., Baia J.E., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
 RA Colauto N.B., Colombo C., Costa F.R., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferto J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nham A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira S.M., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.W., Tsunako M.H.,
 RA Vallada H., Zait M., Meidanis J., Setubal J.C.;
 RA "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RL Nature 406:151-159(2000).
 DR EMBL; AB004023; AAF84868.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 94 AA; 10032 MW; C1977509FC59C134 CRC64;

Query Match 1.6%; Score 7; DB 16; Length 94;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 YVTALSS 33
 Db 63 YVTALSS 69

RESULT 30
 092437 PRELIMINARY; PRT; 99 AA.
 AC 092437;
 DT 01-NOV-1998 (T-EMBLrel. 08, Created)
 DT 01-NOV-1998 (T-EMBLrel. 08, last sequence update)
 DT 01-DEC-2001 (T-EMBLrel. 19, last annotation update)
 DE ACMPNV orf73.
 GN ORF 59.
 OS Bombyx mori nuclear polyhedrosis virus (BmNPV).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 OC Nucleopolyhedrovirus.
 ON NCB1_TaxID=10458;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=T3;
 RX MEDLINE=97329351; PubMed=9185864;
 RA Kamita S.G., Maeda S.;
 RA "Sequencing of the putative DNA helicase-encoding gene of the Bombyx mori nuclear polyhedrosis virus and fine-mapping of a region involved in host range expansion.";

RL Gene 190:173-179(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=T3;
 RX MEDLINE=99281911; PubMed=10355780;
 RA Gomi S., Majima K., Maeda S.;
 RT "Sequence analysis of the genome of Bombyx mori nucleopolyhedrovirus.";
 RT nucleopolyhedrovirus.";
 RL J. Gen. Virol. 80:1323-1337(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=T3;
 RA Maeda S.;
 RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=T3;
 RA Gomi S.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; L33180; AAC63745.1; -;
 SQ SEQUENCE 99 AA; 11486 MW; 83BBBEF76B0E51E7 CRC64;

Query Match 1.6%; Score 7; DB 12; Length 99;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 KKLISN 241
 Db 57 KKLISN 63

RESULT 31
 08B9H9 PRELIMINARY; PRT; 99 AA.
 AC 08B9H9;
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, last annotation update)
 DE Hypothetical protein.
 OS Rachiplusia ou multiple nucleopolyhedrovirus.
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 OC Nucleopolyhedrovirus.
 ON NCB1_TaxID=80366;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Boming B.C., Harrison R.L.;
 RT "The Rachiplusia ou multiple nucleopolyhedrovirus genome sequence.";
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY145471; AAN28138.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 99 AA; 11580 MW; 457D81B8A5A33B5 CRC64;

Query Match 1.6%; Score 7; DB 12; Length 99;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 KKLISN 241
 Db 57 KKLISN 63

RESULT 32
 P82726 PRELIMINARY; PRT; 102 AA.
 ID P82726
 AC P82726;
 DT 01-OCT-2002 (T-EMBLrel. 22, Created)
 DT 01-OCT-2002 (T-EMBLrel. 22, last sequence update)
 DT 01-OCT-2002 (T-EMBLrel. 22, last annotation update)
 DE Hypothetical protein LCR11 precursor.
 GN LCR11.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucoside II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Scheller C., Wambutt R., Murphy G., Volckaert G.,
 Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
 Harris B., Ansgore W., Brandt P., Grivell L.A., Rieger M.,
 Weichselgattner M., de Simone V., Obermaier B., Macho R., Mueller M.,
 Kreis M., Deleney M., Puigdomenech P., Watson M., Schmidheini T.,
 Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 Vos P., Honeisel U., Zimmermann W., Wedler H., Ridley P.,
 Langham S.-A., McCullagh B., Bilham L., Robben J.,
 Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbussche F.,
 Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 Weizengger T., Bothé G., Rampsberger U., Hilbert H., Braun M.,
 Holter E., Brandt A., Peters S., Van Scaveren M., Dirke W.,
 Woolfman P., Klein Lankhorst R., Rose M., Hauf J., Koeltter P.,
 Bernier S., Hempel S., Feldpausch M., Lambirth S., Van den Daele H.,
 De Keyser A., Buysaert C., Gielen J., Villarroel R., De Clercq R.,
 Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 Clark L., Doggett J., Hall S., Kay M., Leonard N., McElay K., Mayes R.,
 Petre A., Kajandream M.A., Lyne M., Benes V., Rechmann S.,
 Bortova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
 Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 Gabel C., Fuchs M., Fartmann B., Grandeth K., Danner D., Herzl A.,
 Neumann S., Argitrou A., Vitale D., Liguori R., Piravandi E.,
 Masene O., Quigley F., Clabaud G., Muendlein A., Felber R.,
 Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 Chedors F., Cooke R., Berger C., Montfort A., Casachubeta E.,
 Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
 Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 Heinen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
 Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 Parrell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 Antonou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 Swaby I.K., O'Shaughnessy A., Rodriguez W., Holtman U., Tili S.,
 Gracat S., Shohdy N., Hasegawa A., Hamed A., Lochi M., Johnson A.,
 Chen E., Maria M., Martensen R., McCombie W.R.,
 "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana."
 RT thaliana."
 RL Nature 402:769-777 (1999).
 RN [2]
 RP IDENTIFICATION.
 RX PubMed=11437247;
 RA Vanoschuyse V., Mies C., Dumas C., Cock J.M.,
 RT "Two large Arabidopsis thaliana gene families are homologous to the
 RT Brassica gene superfamily that encodes pollen coat proteins and the
 RT male component of the self-incompatibility response."
 RL Plant Mol. Biol. 46:17-34 (2001).
 DR EMBL; AL050399; -; NOT ANNOTATED_CDS.
 KW Hypothetical protein; Signal.
 FT SIGNAL
 FT CHAIN 1 29
 FT CHAIN 30 102
 SQ SEQUENCE 102 AA; 10897 MW; 518129BC94E8EAL CRC64;
 POTENTIAL.
 HYPOTHETICAL PROTEIN LCRI1.
 Query Match 1.6%; Score 7; DB 10; Length 102;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 33
 ID Q98WV2 PRELIMINARY; PRT; 103 AA.
 AC Q98WV2;
 DT 01-JUN-2001 (TRENBLREL. 17, Created)
 DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)
 DT 01-OCT-2002 (TRENBLREL. 22, Last annotation update)
 DE Envelope glycoprotein (fragment).
 GN ENV.
 OS Human immunodeficiency virus 1.
 OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=4492;
 RX MEDLINE=21134664; PubMed=11242148;
 RA Thomson M.M., Delgado E., Manjon N., Ocampo A., Villahermosa M.L.,
 Martin A., Herrero I., Cuevas M.T., Vazquez-de Parga E.,
 Perez-Alvarez L., Medrano L., Rabada J.A., Najera R.,
 RT "HIV-1 genetic diversity in Galicia, Spain: BG intersubtype
 RT recombinant viruses are circulating among injecting drug users."
 RL AIDS 15:509-516 (2001).
 DR EMBL; AY017420; AAK14720.1;
 DR InterPro; IPR000777; GP120.
 DR Pfam; PF00516; GP120.1.
 KW AIDS; Coat protein; Glycoprotein.
 FT NON_TER 1
 FT TER 103
 SQ SEQUENCE 103 AA; 11416 MW; 79819EB29A4A9CEE CRC64;
 Query Match 1.6%; Score 7; DB 15; Length 103;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 34
 ID Q9UX9 PRELIMINARY; PRT; 110 AA.
 AC Q9UX9;
 DT 01-MAY-2000 (TRENBLREL. 13, Created)
 DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE Y62E10A.1 protein.
 GN Y62E10A.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lloyd C.R.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018 (1998).
 DR EMBL; AL132865; CAB60595.1;
 DR WormPep; Y62E10A.1; CE22694.
 DR InterPro; IPR001813; 60s ribosomal
 DR InterPro; IPR001859; Ribosomal P2.
 DR Pfam; PF00428; 60s_Ribosomal_P2.
 DR PRINTS; PRO0456; RIBOSOMALP2.
 SQ SEQUENCE 110 AA; 10871 MW; E3DC2B5537A7116C CRC64;

Query Match 1.6%; Score 7; DB 5; Length 110;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 SAGLVSV 156
|||||
59 SAGLVSV 65

RESULT 35

Q9E487 PRELIMINARY; PRT; 113 AA.
ID Q9E487
AC Q9E487;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OC NCBI_Taxid=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20490814; PubMed=11032864;
RA van Rij R.P., Blaak H., Visser J.A., Brouwer M., Rientema R.,
RA Broersen S., de Roda Husman A.M., Schutemaker H.;
RT "Differential coreceptor expression allows for independent evolution
of non-synctium-inducing and syncytium-inducing HIV-1.";
RT J. Clin. Invest. 106:1039-1052(2000).
RL EMBL; AF258985; AAG25026.1; -
DR InterPro: IPR000777; GP120.
DR Pfam: PF00516; GP120; 1.
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1
FT NON_TER 113
SQ SEQUENCE 113 AA; 12684 MW; 2FA353C237FD0D5E CRC64;

Query Match 1.6%; Score 7; DB 15; Length 113;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 KNIRKGL 340
|||||
37 KNIRKGL 43

RESULT 36

Q9E488 PRELIMINARY; PRT; 113 AA.
ID Q9E488
AC Q9E488;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OC NCBI_Taxid=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20490814; PubMed=11032864;
RA van Rij R.P., Blaak H., Visser J.A., Brouwer M., Rientema R.,
RA Broersen S., de Roda Husman A.M., Schutemaker H.;
RT "Differential coreceptor expression allows for independent evolution
of non-synctium-inducing and syncytium-inducing HIV-1.";
RT J. Clin. Invest. 106:1039-1052(2000).
RL EMBL; AF258985; AAG25026.1; -
DR InterPro: IPR000777; GP120.
DR Pfam: PF00516; GP120; 1.
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1
FT NON_TER 113
SQ SEQUENCE 113 AA; 12717 MW; 9A9D62F9BD5430F8 CRC64;

Query Match 1.6%; Score 7; DB 15; Length 113;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 KNIRKGL 340
|||||
37 KNIRKGL 43

RESULT 37

Q9E491 PRELIMINARY; PRT; 113 AA.
ID Q9E491
AC Q9E491;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OC NCBI_Taxid=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20490814; PubMed=11032864;
RA van Rij R.P., Blaak H., Visser J.A., Brouwer M., Rientema R.,
RA Broersen S., de Roda Husman A.M., Schutemaker H.;
RT "Differential coreceptor expression allows for independent evolution
of non-synctium-inducing and syncytium-inducing HIV-1.";
RT J. Clin. Invest. 106:1039-1052(2000).
RL EMBL; AF258981; AAG25022.1; -
DR InterPro: IPR000777; GP120.
DR Pfam: PF00516; GP120; 1.
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1
FT NON_TER 113
SQ SEQUENCE 113 AA; 12731 MW; 9A9D74BAC97530F8 CRC64;

Query Match 1.6%; Score 7; DB 15; Length 113;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 KNIRKGL 340
|||||
37 KNIRKGL 43

RESULT 38

Q9B8A4 PRELIMINARY; PRT; 116 AA.
ID Q9B8A4
AC Q9B8A4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NADH dehydrogenase subunit 3.
GN NAD3.
OS Trichinella spiralis.
OC Trichinellidae.
OC Eukaryota; Metazoa; Nematoda; Enoplea; Trichocephalida;
OC Trichinellidae; Trichinella.
OC NCBI_Taxid=6334;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21100340; PubMed=11156984;
RA Lavoie D.V., Brown W.M.;
RT "Trichinella spiralis mtDNA. A nematode mitochondrial genome that
encodes a putative ATP8 and normally structured tRNAs and has a gene
arrangement reliable to those of coelomate metazoans.";
RT Genetics 157:621-637(2001).
RL [2]
RP SEQUENCE FROM N.A.
RX Lavoie D.V., Brown W.M.;
Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF293969; AAK12904.1; -.
DR InterPro: IPR000440; Oxidored_q4.
DR Pfam: PF00507; oxidored_q4; 1.
KW Mitochondrion.
SQ SEQUENCE 116 AA; 13629 MW; C3CEE2A0BA078BCE CRC64;

Query Match 1.6%; Score 7; DB 8; Length 116;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 FDLFIIT 108
Db 67 FDLFIIT 73

RESULT 39
O8LCV8 PRELIMINARY; PRT; 124 AA.
AC O8LCV8;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein. (Mouse-ear cress).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosida II; Brassicales; Brassicaceae; Arabidopsie.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volkovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation.";
RL Genome Biol. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: A0086382; AAM64449.1; -.
KW Hypothetical protein.
SQ SEQUENCE 124 AA; 13779 MW; 496A29F73425F53 CRC64;

Query Match 1.6%; Score 7; DB 10; Length 124;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 157 VATSILS 163
Db 76 VATSILS 82

RESULT 40
O84225 PRELIMINARY; PRT; 129 AA.
AC O84225;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CHLTR hypothetical protein.
GN CT222.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxId=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/Cx;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;

RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
DR EMBL: AB001296; AAC67814.1; -.
KW Complete proteome.
SQ SEQUENCE 129 AA; 13931 MW; EA7B8BF971327B1B CRC64;

Query Match 1.6%; Score 7; DB 16; Length 129;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 147 LGSSAGL 153
Db 63 LGSSAGL 69

Search completed: August 8, 2003, 09:52:40
Job time : 103 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 15, 2003, 14:03:28 ; Search time 2814 Seconds
(without alignments)
3731.178 Million cell updates/sec

Title: US-10-069-062-7
Perfect score: 2230
Sequence: 1 MSKAFSAPGKAFLAGYLV.....DLBQTEGVLEKEDYIGL 432

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 segs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/sgn2.1/USPTO.spool/US10069062/runat_07082003.161503.4079/app.query.fasta_1.583
-DB=EST -QFMT=fastap -SUFIX=1st -MINMATCH=0.1 -DOOPCL=0 -DOOEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biom62 -TRANS=human40.cdi -LIST=45
-DOCALLIG=200 -THR SCORE=Pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10069062 @CGN 1.1.2810 @runat_07082003.161503.4079 -NCPU=6 -ICPU=3
-NO MAP -IARGOUEXY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEX=7

Database :
1: EST:*
2: em_estba:*
3: em_esthum:*
4: em_estcun:*
5: em_estcov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_estc1:*
10: gb_estc2:*
11: gb_hc:*
12: gb_estc3:*
13: gb_estc4:*
14: gb_estc5:*
15: em_estcun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_ylt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vr1:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	701.5	31.5	920	29	CNS06WZT
2	603.5	27.1	1059	29	CNS06WZT
3	397	17.8	917	29	CNS06WZT
4	320	14.3	753	14	CB971269
5	316	14.2	778	12	B1749876
6	300.5	13.5	769	12	EM076384
7	281	12.6	844	13	BQ118666
8	280.5	12.6	689	13	B0635656
9	264	11.8	494	10	BF596003
10	245	11.0	745	13	B0870439
11	242	10.9	703	13	B0862174
12	241	10.8	716	13	B0866497
13	239	10.7	625	14	CD573486
14	238	10.6	715	13	B0860948
15	236	10.5	482	29	B2304881
16	234.5	10.4	728	9	A1727861
17	231	10.4	728	13	B0870484
18	226	10.1	752	10	BG585318
19	223	10.0	515	14	CD573451
20	222	10.0	705	13	B0869724
21	222	10.0	712	28	AQ911415
22	219.5	9.8	717	12	B0343297
23	216.5	9.7	775	13	B0004622
24	212	9.5	584	12	BM520154
25	212	9.5	664	13	B0861999
26	208.5	9.3	877	10	BF272204
27	208.5	9.3	600	13	B0049350
28	207.5	9.2	594	28	AQ911417
29	204.5	9.2	612	12	BM276808
30	204.5	9.2	608	28	AQ911416
31	198.5	8.9	549	12	B1972859
32	194.5	8.7	817	29	B2701281
33	182.5	8.2	817	29	BF070746
34	180.5	8.1	750	13	BQ110265
35	180.5	8.1	750	13	BQ110265
36	180	8.1	639	13	B0869223
37	179.5	8.0	617	10	BG595511
38	176.5	7.9	513	10	BE494575
39	173	7.8	489	13	B0588169
40	173	7.8	515	14	CA684704
41	173	7.8	973	29	CNS07AVQ
42	172.5	7.7	488	28	AZ931444
43	171	7.7	405	12	B1187231
44	170	7.6	624	13	BQ861358
45	169.5	7.6	641	10	BG448990

ALIGNMENTS

RESULT 1	CNS06WZT	920 bp	DNA	linear	GSS 06-JUL-2001
LOCUS	CNS06WZT				
DEFINITION	T3 end of clone AX0A036H05 of library AX0A from strain CBS 7064				
ACCESSION	AL419103				
VERSION	AL419103.1				
KEYWORDS	GI:12202281				
SOURCE	GSS				
ORGANISM	Pichia farinosa				
	Pichia farinosa				
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
	Saccharomycetales; Saccharomycetaceae; Pichia.				
REFERENCE	1 (bases 1 to 920)				

AUTHORS Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brothier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Winkler, P. and Weissenbach, J.

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of Yeast species for molecular evolution studies

JOURNAL FEMS Lett. 487 (1), 3-12 (2000)

MEDLINE 20584711

PUBMED 11152876

REFERENCE 2 (bases 1 to 920)

AUTHORS de Montigny, J., Spelner, C., Souciet, J., Tekala, F., Dujon, B., Winkler, P., Artiguenave, F. and Potier, S.

TITLE Genomic exploration of the hemiascomycetous yeasts: 15. *Pichia sorbitophila*

JOURNAL FEMS Lett. 487 (1), 87-90 (2000)

MEDLINE 20584725

PUBMED 11152890

REFERENCE 3 (bases 1 to 920)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage, 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

source 1..920
/organism="Pichia farinosa"
/mol_type="genomic DNA"
/strain="CBS 7064"
/db_xref="taxon:4920"
/clone="AX0A036H05"
/clone_1ib="AX0AA"
/note="end : T3"
3..>893
/note="similar to *Saccharomyces cerevisiae* ORF YMR220w [ERG8 ; phosphomevalonate kinase]"
/evidence="not_experimental"

BASE COUNT 219 a 251 c 248 g 201 t 1 others

ORIGIN

misc_feature <3..>893
/note="similar to *Saccharomyces cerevisiae* ORF YMR220w [ERG8 ; phosphomevalonate kinase]"
/evidence="not_experimental"

Alignment Scores:

Pred. No.: 9.7e-68 **Length:** 920

Score: 701.50 **Matches:** 148

Percent Similarity: 65.67% **Conservative:** 49

Best Local Similarity: 49.33% **Mismatches:** 80

Query Match: 31.46% **Indels:** 23

DB: 29 **Gaps:** 7

US-10-069-062-7 (1-432) x CNS06WZT (1-920)

QY 113 GYTHHSERGLNGIULASPTHTGTHLSTHRSERASNGIYGLUYSHTHRPPE 132

DB 3 GGGTACCATCTACAGATACACGACGACCCCTGTCTACAGACGACGCGCATTC 62

QY 133 LEUYTHHSERARALALETHRGUVALGULYSTHGLYLEUGLYSERSERALAGLY 152

DB 63 CTATACCATTAAGAGAAATCAACGAGGTTGCAGAAAACGAGACTTGCGGCGGGA 122

QY 153 LEUVALSERVALVALATHRSERLEUUSERRHSPHEILEPROASN---VALLESER 171

DB 123 TTGGTGCTAGTAGTAACGCGGCGCATTCCTCTCTACTTCATCTGCGACGCGCATTTAT 182

QY 172 THRLEUYSAPLLEUHLLEUHLASNAVALAAGLNULEALHLSYTYRALAGLNUYSYS 191

DB 133 TCACCTCCCAACACGCGTCTCATATACGACGAGNAGTCTACTGTACAGCCCAAGAG 242

QY 192 ILEGLYSERGLYPHEASPVALLATHRALILETYRGLYLEUULEVALTYRARGHPHE 211

DB 243 ATCGGCTCGGTTTGCATGTTGCAGACAGCTGTGACGCTCATGTGTACCGCGTTTC 302

QY 212 GINPROALALEUHLLEUHLASPVALL-----PHEGINVALLEUGLUSERSAPRO 227

DB 303 CAGCTTACTTGTATGACGAAATGACTTAAGACAGACGATTCACAGTAAAGAGTACTCT 362

QY 228 GLU-----LYPHEPROTHRGLEUULYSLEULEUGLUSERSAPROGLU 244

DB 353 ACGCTAGCGCTTGACTACAAAGACTCTTACAGACGATAGGCGTTGACTGCGATTC 422

QY 245 LYSHISGLUARGYVETHLEUPROTYRGLYILEYSLLEULEMETGLYASPVALLYSGLY 264

DB 423 AAAATAGAAATATGCGCTTCCGCGCGCTGACACTCGTAATGGAGAGCGTCTGCTTC 482

QY 265 GLYSERGLUHPROLYLEUVALSERARGVALLEUGINTPLYSGLULYSERGLU 284

DB 483 GGTACGCGACGCCCAATGTGATCGCTAGTCTCAAGTGAAGACGAAACCTCTGTC 542

QY 285 GLUSERSERVALVALTYRASPGLNLEUHLASNAVALASN-----LEU 298

DB 543 GAGGCGGAGAAACTATTGAGAGCTTAATAAGGCATATGTTGCTCATCGCGGCATTA 602

QY 299 GINPHECTLYSGULLEUARGLUMETARGLU---LYSTYRASPSEASAPROGLUTHR 317

DB 603 TCACCTGCTTCAAGAACTTCACGCGCGGCGGAGGAGGCGCTACAGGCTACCTCAAGAC 662

QY 318 TYRILEYSGULLEUASP-----HISSEVALGLUPROLEU----- 329

DB 663 TACAGGTTGCAATGAAAGGTTGCGCGCGGAAATGCTGAAGGAAACCGTTGCGACGGTT 722

QY 330 THRVALALALELTYASNLLEARGLYSGLYLEUGLNULEATHRGINLUSERSERGLUVAL 349

DB 723 TCCGAGGACCATCGCACGCGTGGCGGACCACTTGGCGGCTTACAGATCGCAAGCGGCC 782

QY 350 PROILEGLUPROASPVALLGINTHGLNLEULEUHLASPVARGYSGINLNULEPROGLYCY 369

DB 783 GATATCGAACACCATCTCAAGACAGCCCTCTTGACAGATGACATGCTCTCCGCGCTT 842

QY 370 VALGLYGLYVALVALPROGLVALAGLYGLYTYRASPVALLEALVALLEUVALLEUGLU 389

DB 843 CTGGCGGTTGTGTCTCGCGCGGCTTGCTTCAAGCGCATCTGCTTCTGAG 902

RESULT 2

CNS06MRJ/c 1059 bp DNA linear GSS 17-JUN-2001

LOCUS T3 end of clone AU0A006F09 of library AU0A from strain CBS 3082

DEFINITION OF *Saccharomyces kluyveri*, genomic survey sequence.

ACCESSION AL405845

VERSION AL405845.1 GI:12169063

KEYWORDS GSS.

SOURCE *Saccharomyces kluyveri*

ORGANISM *Saccharomyces kluyveri*

REFERENCE Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brothier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Winkler, P. and Weissenbach, J.

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of Yeast species for molecular evolution studies

JOURNAL FEMS Lett. 487 (1), 3-12 (2000)

MEDLINE 20584711

PUBMED 11152876

REFERENCE 2 (bases 1 to 1059)

AUTHORS Neuveglise, C., Bon, E., Lepingle, A., Wincker, P., Attiguenave, F., Galliard, C. and Casaregola, S.
 TITLE Genomic exploration of the hemiascomycetous yeasts: 9.
 JOURNAL Saccharomyces kluyveri
 MEDLINE FEMS Lett. 487 (1), 56-60 (2000)
 PUBMED 20584719
 REFERENCE 11152884
 3 (bases 1 to 1059)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces marxianus var. marxianus, Kluyveromyces lactis var. lactis, Kluyveromyces fragilis var. fragilis, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
 FEATURES
 SOURCE
 1. 1059
 /organism="Saccharomyces kluyveri"
 /mol_type="genomic DNA"
 /strain="CBS 3082"
 /db_xref="taxon:4934"
 /clone="AUDA006F09"
 /clone_11b="AUDA00A"
 /note="end : T3"
 complement (<6..>824)
 ERG8 ; phosphomethyltransferase kinase]"
 /evidence="not_experimental"
 misc_feature
 214 c 211 g 300 t 2 others
 BASE COUNT 332 a 214 c 211 g 300 t 2 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.16e-56 Length: 1059
 Score: 603.50 Matches: 133
 Percent Similarity: 66.18% Conservative: 49
 Best Local Similarity: 48.36% Mismatches: 80
 Query Match: 27.06% Indels: 13
 Gaps: 7
 DB: 29
 US-10-069-062-7 (1-432) x CNS06MRJ (1-1059)
 Oy 2 SerlysalapheserAlaProGlyYlysalapheluaiaGlyGlyTyrLeuValleuGlu 21
 Db 815 TCCAGAGATTCAGTGCACGACGTAAGCCCTTACGTGATGATCTTCTCTAGAT 756
 Oy 22 ProiletyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrPro 41
 Db 755 CCCAAATATATAGATCTTATGCTGCTGCTTCACTGATGATCTTCTGCTGATCTCAG 696
 Oy 42 LysGlyThrSerLeuLysGluSer---ArgIleLysIleSerSerProGlnPheAlaAsn 60
 Db 695 AATGATTCGAAAGATTAATCAAGATCGAGTTCAGTTCAGTCTCAATTCAGTAA 636
 Oy 61 GlyGluTyrGluTyrHisIleSer---SerAnthrGluLysProArgGluValGlnSer 79
 Db 635 GACTGCTGCTGCTATGACTGATGATGATGATGATGATGATGATGATGATGATGAT 576
 Oy 80 ArgIleAsnProPheLeuGluAlaThrIlePheIleValleuAlaTyrIleGlnProThr 99
 Db 575 CTGAGAAATCATTCATGCTGAAAGACATGCTGAAACGCTTGAATCTTCTGAGCCCAA 516
 Oy 100 GluAlaPhe-----AspLeuGluIleIleIleIleTyrSerAspProGlyTyrHisSerGln 117
 Db 515 TCTGACATGACTCCAAATATTTTCATGCAAGATTCATTCATCTGCTATCTCA 456

Oy 118 GluAspThrGluThrLysThrSerSerAnthrGluLysThrPheLeuTyrHisSerArg 137
 Db 455 GAGGAGAGTGTCCCGAG-----AGGAATAGTTTCAAAACATTTAATTTCCATGAAC 402
 Oy 138 AlaIleThrGluValGluLysThrGlyLeuGlySerSerAlaGlyLeuValSerValVal 157
 Db 401 ACCATCAACAAGTCCCAAGATCGGCTGGCTCTTCTGCGGATGTTGAAGTTCTA 342
 Oy 158 AlaThrSerLeuLeuSerHisPheIleProAnthrValIleSerThrAnthrAsp----- 175
 Db 341 AGTCTGTTTAACTCATCTGATCTTCAACCTGATCTTATGATGATCTTCCAGACCATTTG 282
 Oy 176 ---IleLeuHisAsnValAlaGlnIleAlaHisCysTyrAlaGlnLysValIleGlySer 194
 Db 281 ACTTAATTCATATACATGACCAAGTTCTCATGCTGAGTCAAGGTAAGTGGTATGC 222
 Oy 195 GlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyrArgArgPheGlnProAla 214
 Db 221 GGATTCGATGTTGGGACGACCATTTGGATCTTATGTTATCAGAGATTCGCCGACGAA 162
 Oy 215 LeuIleAsnAspValPheGlnValLeuGluSerAspProGluLysPheProThrGluLeu 234
 Db 161 CTGATTAATTAAT-----TTACCACTGCTGATCTGCTGATTAACACGATTCCTC 111
 Oy 235 LysLysLeuIle---GluSerAnthrGluGluLysHisGluArgCysThrLeuProTyr 253
 Db 110 GTGAAGTTGTTGATGAATGATGATGATGATGATGATGATGATGATGATGATGAT 51
 Oy 254 GlyIleLysLeuLeuMetGlyAspValLysGlyGlySerGluThr 268
 Db 50 GCGTTAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6
 RESULT 3
 CNS06MRJ 917 bp DNA linear GSS 06-JUL-2001
 LOCUS T3 end of clone AX0A034D03 of library AX0A from strain CBS 7064
 DEFINITION of Pichia farinosa, genomic survey sequence.
 ACCESSION AL418769
 VERSION AL418769.1 GI:12201829
 SOURCE GSS.
 ORGANISM Pichia farinosa
 Pichia farinosa
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Pichia.
 REFERENCE
 1 (bases 1 to 917)
 Souciet, J.L., Aigle, M., Attiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brotter, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Florent, B., Malpertuy, A., Neuveglise, C., Ozier-Kalopoulou, O., Potier, S., Saurin, M., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P. and Weissbach, J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
 FEMS Lett. 487 (1), 3-12 (2000)
 2 (bases 1 to 917)
 de Montigny, J., Spehner, C., Souciet, J., Tekala, F., Dujon, B., Wincker, P., Attiguenave, F. and Potier, S.
 Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila
 FEMS Lett. 487 (1), 87-90 (2000)
 JOURNAL
 MEDLINE 20584725
 PUBMED 11152890
 3 (bases 1 to 917)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,

Qy 220 PheGlnValLeuGluSerAspProGluLysPheProThrGluLeuValLeuGlu 239
 Db 122 CAGGTTCAGTGAAGAAGACTGCC-----ATACAAAGAAGCATCATGATATATAA 175
 Qy 240 SerAspTyrGluGluLysHisGluArgCysThrLeuProTyrGlyTyrLeuLeuMet 259
 Db 176 GGAATTTGGGACCATGAAAGAGCTAAATTTTCCTTCCCTGACCGTATGATCACTTTGGCTA 235
 Qy 260 GlyAspVal---LysGlyGlySerGluThrProLysValSerValSerValLeuGlu 278
 Db 236 GGAAGAACAGGCGCTGGAGGTTTCATCAACCATATGATGAGTGGCTGTTAAAGAGTGG 295
 Qy 279 LysIleGluLysProGluGluSerSerValValTyrAspGlnLeuAsnSerValLeu 298
 Db 296 CAAAGGTTCGACCTGAGAAAGCTCTAGAAACATGAGAAAGTTGTCAGCGCAATTC 355
 Qy 299 GlnPheMetLysGluLeuArgGluMetArgGluLysTyrAspSerAspProGluThrTyr 318
 Db 356 GCACCTGGAACCAATTCAGATTTTGAATTAATTAGCAGAGAAATTTGGAGATGCATAT 415
 Qy 319 -----LysGluLeuAspHisSerVal 326
 Db 416 AATGTGTGATCAACAGTTCAGCATTCATAGTCAGAGAGTGGATGAGCAAGCCACT 475
 Qy 327 GlnProLeuThrValAlaIleLysAsn----- 335
 Db 476 GGGCGACAGCAAGAAGAGCTGTTAATGTCATTTAGCAGCAAGATGCATGCTTAG 535
 Qy 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspVal 355
 Db 536 ATCAGAAATTAATATGCGCAGATGGGCTGAGCTGAGCATCCGATGAACCTGATCT 595
 Qy 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
 Db 596 CAACCTCGGCTTTTGATGCTACTATGAGCAATGAGAGATTCATATGGCTGGAGTTCCA 655
 Qy 376 GlyAlaGlyGlyTyrAspAlaIleAlaValLeuValLeuGluAsnGlnValGlyAsn 394
 Db 656 GGACAGAGGTGATTTGATGCAATTTTTCCTGCTCACTTACGAGGATTCAGATGACAC 712

RESULT 5
 B1749876 778 bp mRNA linear EST 25-SEP-2001
 LOCUS Fg02_01b06_R Fg02_AAFRC Fusarium graminearum mycelium
 DEFINITION Giberella zeae cDNA clone Fg02_01b06, mRNA sequence.
 B1749876
 ACCESSION B1749876.1 GI:15771678
 VERSION
 KEYWORDS
 SOURCE EST
 ORGANISM Giberella zeae
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Nectriaceae; Giberella.
 1 (bases 1 to 778)
 HARRIS,L.J., GLASSCO,T., ROCHELEAU,H., ALLARD,S., CHAPADOS,J.,
 CONROUX,P., DE MOORS,A., HATTORI,J.T., ONEILL,T., ROBERT,L.S.,
 SINGH,J.A., SPRETT,D. and TINKER,N.A.
 Expressed Sequence Tags from Fusarium graminearum mycelium
 JOURNAL Unpublished
 COMMENT Contact: Harris, Linda J.
 Eastern Cereal and Oilseed Research Centre
 Agriculture and Agri-Food Canada
 Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
 CANADA
 Tel: (613) 759-1314
 Fax: (613) 759-6566
 Email: harris.lj@em.agr.ca.
 FEATURES
 source 1..778
 /organism="Giberella zeae"
 /mol_type="mRNA"
 /strain="DNOM 180378"
 /db_xref="taxon:5518"
 /clone="Fg02_01b06"

/tissue type="Mycelial tissue"
 /dev_stage="Asexual"
 /lab_host="E. coli (Sure cells)"
 /clone_lib="Fg02_AAFRC Fusarium graminearum mycelium"
 /note="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: EcoRI;
 Site 2: XhoI; Mycelial tissue was collected from V8 agar
 plates after a growth period of 6-7 days at 25 C with 14
 hrs (FR/UV) day light exposure. Mycelia was ground in
 liquid nitrogen prior to its storage at -80 C until RNA
 extraction. Directional cloning with 5' end of cDNA cloned
 into EcoRI site of pBluescript and 3' end of cDNA cloned
 into XhoI site of pBluescript (Stratagene, La Jolla, CA)."
 BASE COUNT 157 a 233 c 177 g 210 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.21e-24 Length: 778
 Score: 316.00 Matches: 84
 Percent Similarity: 54.79% Conservative: 36
 Best Local Similarity: 38.36% Mismatches: 69
 Query Match: 14.17% Indels: 31
 DB: 12 Gaps: 9

US-10-069-062-7 (1-432) x B1749876 (1-778)
 Qy 4 AlaphSerAlaProGlyLysValAlaPheLeuAlaGlyLysValLeuGluProIle 23
 Db 140 GCCCTTCGGCTCTGGCAAGGTTCTCTCGCGGATGTCCTTATGACCAAGAG 199
 Qy 24 TyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIle----- 39
 Db 200 TACACGGCTTTGTGTTCGGCTCAATGCTGCATCAATATCATTCGACAGATATCCAC 259
 Qy 40 ThrProLysGlyThrSerLeuLysGluSerArgIleLysIleSerSerProGlnPheAla 59
 Db 260 ACGACTCGCTGTGTCAGCTCAGCTGAG-----ATCGGTGTATGACCCCTCAGTTCCT 313
 Qy 60 AsnGlyLysThrGlu-----TyrHisIleSerSerAsn-----Thr 71
 Db 314 GATCGCAATGGCGGATATGATATACCTTGTGGCGAGCGCGGCTATCAAGTTACT 373
 Qy 72 GlnLysProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThrIlePheIle 91
 Db 374 CAATGCAAGTCGAGCTCAGATCAACCCCATCTTCGTGACAGAGACTTAGCTAT 433
 Qy 92 ValLeuAlaTyrIle-----GlnProThrGluAlaPhe--AspLeu 104
 Db 434 GCTCTCACTTACATTCGACCGGTGCTGGCACCGCTCTAGTCAACGCTTGGCATCTGCT 493
 Qy 105 GlnIleIleLysSerAspProGlyTyrHisSerGlnGluAspThrGluThr 124
 Db 494 CGCTTATATCTTCCTGGCAATGATTAATCTTCACTCGGAGTGTGACACACTGCC 553
 Qy 125 SerSerAsnGlyLysThrPheLeuTyrHisSerArgAlaIleThrGluValGluLys 144
 Db 554 TCGGGCGCG-----TTGGCAAGTTCCTGTGACCTCGACAGCATGCTAACAAG 601
 Qy 145 ThrGlyLeuGlySerSerAlaGlyLeuValSerValAlaThrSerLeuLeuSerHis 164
 Db 602 ACTGATCTTGATTTCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 661
 Qy 165 PheIlePro---AsnValIleSerThrAsnLysAsp-----IleLeuHisAsn 179
 Db 662 TATCTACCGAGAGATCTCTTATGATGATTCGACAGAGAAAGAGAGAGCTTCAATAC 721
 Qy 180 ValAlaGlnIleAlaHisCysTyrAlaGlnLysLysIleGlySerGlyPheAspVal 198
 Db 722 TTGGCCAGGCT-CTCACTGTGCTGCCCAAGGCAAGCTTGGCTCGGCTTCGATGTC 777

RESULT 6
 BM076384 769 bp mRNA linear EST 05-FEB-2002
 LOCUS TREST-A4106 TREST-A Hypocrea jecorina cDNA clone Tr-A4106 5'
 DEFINITION

XhoI: supplier: Combination of untreated and Phytophthora infestans-created libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes ' tubers, or roots. "

BASE COUNT 229 a 215 c 155 g 245 t
ORIGIN

Alignment Scores:
Pred. No.: 1,15e-20 Length: 844
Score: 281.00 Matches: 84
Percent Similarity: 47.92% Conservative: 54
Best Local Similarity: 29.17% Mismatches: 86
Query Match: 12.60% Indels: 6
Gaps: 10

US-10-069-062-7 (1-432) x BU118666 (1-844)

174 LysAspIle-----LeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
189 GlnLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
770 CAGGGTAAAGTTGGCAGTGGGTTGATGTAGTTCTGCAAGTTATGGAAGTCACGCTAT 711
209 ArgArgPheGlnProAlaLeuIle-----Asn-AspValPheGln----- 221
710 ATCCGGTTTGGCCGAGAGTGTCTCTTCGCCAGAAATGCAAGATGCAACCACTA 651
222 -----ValLeuGlnSerAspProGlnLysPheProThrGlnLeuLysLeuIleG1 239
650 ATAGGAAGTCATGTATGAT-----GTCCTTAA 624
239 uSerAsnThrGlnLysIleGlySerGlyPheProThrGlyLeuLysLeuIleG1 259
623 AGCAAAAGTGGACCACTGAGAGACCAAGTTTTCATGCTCTTAAAGACATTATTA 564
259 tG1LysAspValLys---GlyGlySerGlnThr-ProLysLeuValSerArgValLeuGln 278
563 TGGAGAACCGGGCGTGTGAGAGATCTTCAACCCATCAATGTTGAGCTTTAAGAAT 504
278 rPlLysLysGlnLysProGlnLysSerSerValValTyrAspGlnLeu-----Asn 295
503 GGCAGAGGTCCGACCTCAGATTCCTCAGAAACATGAGAGAGTGTCCAGAGAAACT 444
295 eAlaAsnLeuGlnPheMetLysGlnLeuArgGlnMetArgGlnLysLysSerAsp 315
443 CTGCTCTTGAAGCCGATCTAAATACCTTAAGTAAATTGGCAGAGACATTATCAATGCTT 384
315 roGlnThrTyrIle-----LysGlnLeuAspHisSerV 326
383 ATGAATGATCATCAATGCCCTGCAGCCCTCTCCCTGCAAAAGGTGTGAGAGAGCA 324
326 alGlnProLeuThrVal-----AlaIleLysA 335
323 ATGAACCACTCAAGCAAAATTGTTAAAGATTATGAGACTCGAGATCTATGCTTC 264
335 snIleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGlnValProIleLysProAspV 355
263 GGATTAGGTATCAAGATGCGAAGATGGAGAGATTGCGAATTCCCAATAGAACCAAGAT 204
355 alGlnThrGlnLeuLysAspArgCysGlnGlnIleProGlyCysValGlyValValP 375
203 CGCAATCTCAAGCTCTAGATCTCAATGATATGAGAGAGCTGTGTGGCTGCGCATTC 144
375 roGlyAlaGlyGlyTyrAspAlaIleAlaValLeuValLeuGlnAsnGlnValGlyAsn 395
143 CTGGTGACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 84
395 heLysGlnLys-----ThrLeuGlnAsnProAsp 405
83 TGACTCAAGCTTGAAGTTCACTTAATGTTCTTGAAGTCTAGTAAGAGAGATCTCT-- 28

Qy 405 YrPheHisAsnValTyrTyr 411
Db 27 -----CATGGTGTCTCTGG 13

RESULT 8
BU635656

LOCUS 028D01 Infected Arabidopsis Leaf Arabidopsis thaliana cDNA, mRNA
DEFINITION

ACCESSION BU635656
VERSION BU635656.1 GI:23302911

KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons, core eudicots, rosids

REFERENCE 1 (bases 1 to 689)
Lundsgaard, M., Emmersten, J., Nielsen, K.L., Wilson, I., Somerville, S., and Weindler, K.G.
EST sequencing of Erysiphe cichoracearum infected Arabidopsis plants

AUTHORS Unpublished
CONTACT Karen G. Weindler
INSTITUT FOR BIOTEKNOLOGI
AALBORG UNIVERSITET
SOLGAARDSHOLMVEJ 49, 9000 AALBORG, DENMARK
TEL: +45 96358467
FAX: +45 98141808
EMAIL: kgwebio.au.dk.

FEATURES
source location/qualifiers

1..689
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/dev_stage="Plant 3 weeks old, three days post infection"
/clone_lib="Infected Arabidopsis Leaf"
/note="Organ: Leaf; Vector: pBluescript; Mixed cDNA library of Arabidopsis and E. cichoracearum infected leaf from three weeks old Arabidopsis plants. Plants were harvested 3 days after infection and mRNA oligo dt selected."

BASE COUNT 204 a 123 c 167 g 195 t
ORIGIN

Alignment Scores:
Pred. No.: 9.53e-21 Length: 689
Score: 280.50 Matches: 72
Percent Similarity: 51.09% Conservative: 45
Best Local Similarity: 31.44% Mismatches: 77
Query Match: 12.58% Indels: 35
Gaps: 8

US-10-069-062-7 (1-432) x BU635656 (1-689)

141 GlnValGlnLysThrGlyLeuGlySerSerAlaGlyLeuValSerValAlaIleThrSer 160
10 GAAGTACCAAACTGGCTTAGGTGTTCTTGCAGCAATGCAACAGCTGTGGTTGACGCT 69
161 LeuLeuSerHisPheIleProAsnValIleSer----- 171
70 CTGTTA---CATATCTT---GGAGTGTGACCTATCTGATCATCATGTAAGAGAAAG 123
172 -----ThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
124 TTGGCTGTCTCTATCATGATGTTATTCATATGATAGCAAAAGCTCTCATTTGCTTGA 183
189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
184 CAAAGGAAAGTTCGAGAGTGGGTTGATGATGATGATGATGATGATGATGATGATGAT 243
209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGlnLysSerPro-- 227

```

Db      244 GTTCGCTCTCCAGAAAGCTTGTGCTCAGTTGCTGAGTAAACAGCTCTGCATTA 303
               |||||  |||  : : : : : : : : : : : : : : : : |||||
Qy      228 ---GluyspheProthrGluLeuLysLeuLeuLeuLeuLeuLeuLeuLeuLeu 246
Db      304 AATGAGATTATTGTCATTTTGAAG-----GGAAAATGGACATATAGACA 351
Qy      247 GluArgCysThrLeuProThrGlyLysLeuLeuLeuLeuLeuLeuLeuLeuLeu 265
               : : : : : : : : : : : : : : : : : : : : : : : :
Db      352 ACTGAGCTTCTTACACCACTGATGATCTTTCTTCTGAGAACCTGGAAGTGTGGA 411
Qy      266 SerGluThrProLysLeuValSerArgValLeuGluThrLysLeuLysProGlu 285
               |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      412 TCTCCACACCATCATGATGTGATGATGATGATGATGATGATGATGATGATGATGAT 471
Qy      286 SerSerValValLysArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 305
               : : : : : : : : : : : : : : : : : : : : : : : :
Db      472 GACAGCAAAAGTGGAGAAATTTGTCATGCAAAATTTGAACTGAAACTAAGCTAAC 531
Qy      306 GluMetArgGluLysLysLysLysLysLysLysLysLysLysLysLysLysLys 323
               : : : : : : : : : : : : : : : : : : : : : : : :
Db      532 GATCTGAGCAAAATTAGCTTAAGACCACTGAGATCTTATCTAGATCTTATAGCTTGT 591
Qy      324 -----HisSerValGluProLeuThrValAlaIle 333
Db      592 AGTGTGCTTACTTGAAGAAAGTGGTGTGATGATGATGATGATGATGATGATGAT 651
Qy      334 LysAsnIleArgLysGlyLeuGluAla 342
Db      652 -----ATTAAAGAACTTTAAGAGCA 672

```

RESULT 9
LOCUS BFs96003 494 bp mRNA linear EST 06-DEC-2001
DEFINITION Bu76f09.y1 Gm-c1055 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1055-1026 5' similar to TR:Q9UT88 Q9UT88 PUTATIVE PHOSPHOMETHYLONATE KINASE. ; mRNA sequence.

ACCESSION BFs96003
VERSION BFs96003.1 GI:11688327
KEYWORDS EST
SOURCE Glycine max (soybean)
ORGANISM Glycine max

REFERENCE 1 (bases 1 to 494)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna,A., Bolla,B., Mairra,M., Hallier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished

TITLE Public Soybean EST Project
JOURNAL Contact: Shoemaker R/Public Soybean EST Project
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: esewatson.wustl.edu
This clone is available through: Resgen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: cdueresgen.com
High quality sequence stop: 411.
Location/Qualifiers
1..494
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1055-1026"
/issue_type="Mature seed pods, greenhouse grown"

FEATURES

source

```

/lab host="DH10B"
/clone lib="Gm-c1055"
/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from mature seed pods of greenhouse grown plants prior to
senescence for the cultivar KPI. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with a XhoI restriction site. EcoRI
adapters were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA fragments were
directionally cloned into the EcoRI-XhoI restriction site
of the pBluescript vector. The ligated cDNA fragments
were transformed into DH10B host cells (GibcoBRL). This
library was constructed in the laboratory of Dr. Randy
Shoemaker."

```

BASE COUNT 128 a 96 c 129 g 141 t
ORIGIN
Alignment Scores:
Pred. No.: 4,036-19 Length: 494
Score: 264.00 Matches: 63
Percent Similarity: 57.14% Conservative: 29
Best Local Similarity: 39.13% Mismatches: 45
Query Match: 11.84% Indels: 24
DB: 10 Gaps: 6

US-10-069-062-7 (1-432) x BFs96003 (1-494)

```

Qy      141 GluValGluLysThrGlyLysLeuLysSerValGluValSerValAlaThrSer 160
               |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      33 GAAGTGGCTAAAGCTGTTGGCTCTTCTGACAGCAAGCAATGCTGTAGTGTGCT 92
Qy      161 LeuLeuSerHisPheIleProAsnVal---LleSerThrAsnLys----- 174
               |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      93 CTACTT---CATTACTGATGTTGTAAGCTTCTCTTGAAGATCATCGGAAGG 149
Qy      175 -----AspIleLeuHisAsnValAlaGluIleAlaHisCysTyrAla 188
               |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      150 AAGATGTTGAGATCTTGTATATGTCATTAATAGCTCAACTGACGACATGTATTGCA 209
Qy      189 GluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 208
               |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      210 CAGGGGAAGTGGAGAGGTTGATGTCAGCTCAGCTGTGATGCGAGTCAGCGCTAT 269
Qy      209 ArgArgPheGlnProAlaLeuIleAsnAspVal-----PheGlnValLeuGlu 224
               |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      270 GTGCGTTTTCACGGAAGTATTTCTCCACTAAGCTGCAATATGACAGTGCCTTTA 329
Qy      225 SerAspProGluLysPheProThrGluLeuLysLysLeuLeuLeuLeuLeuLeu 244
               |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      330 TCAGAT-----GGTATCACTGAATATTAAAGAAATTTGGACCAT 371
Qy      245 LysHisGluArgCysThrLeuProThrGlyLysLeuLeuLeuLeuLeuLeuLeuLeu 263
               : : : : : : : : : : : : : : : : : : : : : : : :
Db      372 GACACGACTGATCTCTCTCAACACTTGTATGTCCTTTACTAGAGAACTGTAACT 431
Qy      264 GlyGlySerGluThrProLysLeuValSerArgValLeuGluThrLysLysPro 283
               |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      432 GGTGATATCATCAGCCATGATGATGTTGTCGTTTAAATAATGGCAAAAGCTGACCT 491
Qy      284 Glu 284
               : :
Db      492 CAG 494

```

RESULT 10
LOCUS B0870439 745 bp mRNA linear EST 15-AUG-2002
DEFINITION QGD9A23.yg.ab1 OG_ABCDI lettuce salinas lactuca sativa cDNA clone
ACCESSION B0870439
VERSION B0870439.1 GI:22256848
KEYWORDS EST
SOURCE Lactuca sativa

ORGANISM *Lactuca sativa*
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
 Cichorieae; Lactuca.

REFERENCE
 1 (bases 1 to 745)
 Kozik, A., Michelmore, R. W., Knapp, S., Matvienko, M., Rieseberg, L.,
 Lin, H., van Damme, M., Lavelle, D., Chevaller, P., Ziegler, J., Ellison,
 P., Kolman, J., Slabaugh, M. S., Livingston, K., Zhou, Y., Lai, Z.,
 Church, S., Jackson, L. and Bradford, K.
 Lettuce and Sunflower ESTs from the Compositeae Genome Project
 http://compgenomics.ucdavis.edu/

TITLE
 Unpublished

JOURNAL
 Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Assumundson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig OG_CA_Contig6356, see http://cgdb.ucdavis.edu/
 for details.

FEATURES
 source
 1. 745
 /organism="Lactuca sativa"
 /mol_type="mRNA"
 /cullivar="Salinas"
 /db_xref="taxon:4236"
 /clone="QSD9A23"
 /lab_host="E.coli"
 /clone_1lb="QG ABCDI lettuce salinas"
 /note="Vector: pBRCDNA5f1AB: The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgdb.ucdavis.edu/
 TAG_L1B=QG_ABCDI lettuce salinas
 TAG_TISSUE=chemical induction
 TAG_SEQ=GTGAGCCGCGG"

BASE COUNT 216 a 142 c 152 g 235 t

ORIGIN

Alignment Scores:
 Pred. No.: 1.02e-16 Length: 745
 Score: 245.00 Matches: 76
 Percent Similarity: 45.53% Conservative: 41
 Best Local Similarity: 29.57% Mismatches: 70
 Query Match: 10.99% Indels: 70
 Gaps: 11

US-10-069-062-7 (1-432) x B0870439 (1-745)

Qy 17 TyrleuValleuGluProIleTyrAspAlaTyrValThrAlaIeuSerSerArgMetHis 36
 Db 28 TATGCAATTGTGAAGCCAAATTATGATANA----- 57
 Qy 37 AlAValIleThrProLyGlyThrSerleuYsgIeuSerArgIleYleYleSerSerPro 56
 Db 58 -----CTAAGCCCAAGATGGCTTGGTCATGAGACAGATTGAAGTTAACTTCTTCT 111
 Qy 57 GluPheAlaenGlyIuTyrGluYrHisIleSerSerAen-----ThrGluIuYsPro 74
 Db 112 CAATGCGCAGAG---GAAGCTATCTACAAATGTCACTAAGCAATTTCCTTCAATGC 168
 Qy 75 ArgGluValGlnSerArgIleasnProPheIeuGluuAlaThrIlePheIleValIeuAla 94
 Db 169 ACTTCAGTGAATCAAGG---AACCTTTGTGGAATATGACGCAATATGCTGTAGCA 225

Qy 95 TyrleuGlnProThr-----Glu 100
 Db 226 GCACATATGATGACATTTGGTAATGACAAAGAGATGACATTCAAAAACGCTCTGTA 285
 Qy 101 AlaPheAspLeuGluIlelle-----IleTyrSerAspProGlyTyrHisSer 116
 Db 286 GGTTTATATATACAAATTTTAGTGGCAATGACTTTATATTC-----TATCGCAT 336
 Qy 117 GluIuYsPheThrGlu----- 121
 Db 337 GAGATTGAGACAGCTGGCTTCACTAACCCCGAATCATGGATCTTACACCTTTT 396
 Qy 122 -----ThrYrThrSerSerAsnGlyGluYrThrPheIuYrHisSerArgAla 138
 Db 397 TCTTCATATCTTGAATTTCTTCAAGTGAGAAAGT-----TGT 435
 Qy 139 IleThrGluValGluYrThrGlyLeuGlySerSerAlaGlyLeuValSerValAla 158
 Db 436 AAGCTGAAAGTTGCAAAAGCTGGTGGATCTTCTGACGCAATGACACAGAGTGT 495
 Qy 159 ThrSerleuSerSerHisPhe---IleProAsnValIleSerThr----- 172
 Db 496 GCTGCTTACTTATTAATCTTGGAGTTGTATATCTTCACTTCTAGTGGGCCCAATAT 555
 Qy 173 -----AsnYsAspIleLeuHisAsnValAlaGlnIleAlaHisCyTyrAlaGlnYs 190
 Db 556 CAAGAAATCTGTATGTGGTGCATATATAGCCAAATGGACACTGTATGCTCAAGGA 615
 Qy 191 LysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyrArg 210
 Db 616 AAATGTGGAGTGAATGTGATGTAGTTCAGCTGTTATGGAAGTCAATCGTTATGAGA 675
 Qy 211 PheGlnProAlaLeuIleAsnAspValPheGlnValIeuGluSerAspPro 227
 Db 676 TTTCACCAAGATTATCTTCTGCTCAGAAATATACAGAGCCTTACCA 726

RESULT 11
 B0862174 703 bp mRNA linear EST 14-AUG-2002
 LOCUS OGC20P15.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone
 DEFINITION OGC20P15, mRNA sequence.
 ACCESSION B0862174
 VERSION B0862174.1 GI:22247629
 KEYWORDS EST.

SOURCE
 Lactuca sativa
 ORGANISM
 Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
 Cichorieae; Lactuca.

REFERENCE
 1 (bases 1 to 703)
 Kozik, A., Michelmore, R. W., Knapp, S., Matvienko, M., Rieseberg, L.,
 Lin, H., van Damme, M., Lavelle, D., Chevaller, P., Ziegler, J., Ellison,
 P., Kolman, J., Slabaugh, M. S., Livingston, K., Zhou, Y., Lai, Z.,
 Church, S., Jackson, L. and Bradford, K.
 Lettuce and Sunflower ESTs from the Compositeae Genome Project
 http://compgenomics.ucdavis.edu/

TITLE
 Unpublished

JOURNAL
 Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Assumundson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig OG_CA_Contig6356, see http://cgdb.ucdavis.edu/
 for details.

FEATURES
 source
 1. 703
 /organism="Lactuca sativa"
 /mol_type="mRNA"
 /cullivar="Salinas"

Accession	Version	Keywords	Source	Organism	Reference Authors	Journal Comment	Features
1	1	lab	Zea mays	CDNA, mRNA sequence.			
CD573123							
1	1	GI:31664398					
EST.							
Zea mays							
Zea mays							
Eukaryote:							
Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;							
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD							
clade: Panicoidae; Andropogonaceae; Zea.							
1 (bases 1 to 529)							
Walbot, V.							
Maize ESTs from various cDNA libraries sequenced at Stanford							
University							
Unpublished							
Contact: Walbot V							
Department of Biological Sciences							
Stanford University							
855 California Ave., Palo Alto, CA 94304, USA							
Tel: 650 723 2227							
Fax: 650 725 8221							
Email: walbot@stanford.edu							
Plate: 3529.1.126.1 row: G column: 04.							
Location/Qualifiers							
1..529							
/organism="Zea mays"							
/mol_type="mRNA"							
/cultivar="B73"							
/db_xref="taxon:4577"							
/feature_type="ear"							
/dev_stage="2 mm"							
/lab_host="E. coli XL0R"							
/clone_lib="3529 - 2 mm ear tissue from Schmidt and Hake							
labs"							
/note="Organ: ear; Vector: PAD-GAL4-2.1; Site:1: EcoRI;							
Site:2: XhoI; RNA isolated by Hake lab. 1 million pfu							
amplified. Ampicillin is the selection marker."							
BASE COUNT	145 a	112 c	129 g	143 t			
ORIGIN							
Alignment Scores:							
Pred. No.:	3,68e-16	Length:	529				
Percent Similarity:	51.70%	Matches:	54				
Best Local Similarity:	30.68%	Conservative:	37				
Query Match:	10.67%	Mismatches:	71				
DB:	14	Gaps:	3				
US-10-069-062-7 (1-432) x CD573123 (1-529)							
Oy	154	ValSerValValAlaThrSerLeuLeuSerHisPheIleProAlaValIle-----	1707				
Db	2	ATTCTATTCTTTATGAAAGACAGATTGAAAGCAGTGTCTCCCTCTTACTCCGAGGACA	61				
Oy	171	-----SerThrAsnLysAspIleLeuHisAsnValAlaGlnIle	163				
Db	62	ATCTTCTGGCGCAAAAGGCAAGGAGCAAGCACTTGACCTTGTTCAAGCTATTGGCCAAAGT	121				
Oy	184	AlaHisCysTrpAlaGlnLysLysLysIleGlySerGlyPheAspValAlaThrAlaIleTr	203				
Db	122	GCAACATGTTTACACACAGGAAAATTTGCACTGTTTGAATGTTAGTCAAGCTGCTCAT	161				
Oy	204	GlyLeuIleValLysTrpArgTrpGlnProAlaLeuIleAsnAspValPheGlnValLeu	223				
Db	182	GGAAGTCACAGCTATGATGAGTTTCTCCAGAAATATCTCTCCGCTCAGGCTATTA	238				
Oy	224	GlusSerAspProGlnLysPheProThrGlnLysLysLysLysLysLysLysLysLysLys	243				
Db	239	-----GTTGGAGCTGCTTTCACAGATGATGATGATGATGATGATGATGATGATGAT	222				
Oy	244	GlnLysHisGlnLysTrpLysLeuProTyrGlyLysLysLysLysLysLysLysLysLys	262				
Db	293	CATGAGAAATAAACAGTTTTCATTACTCTCTGATGAGACCTTCTCTGAGGAGACCTG	352				

FEATURES	source
ORIGIN	206 a 135 c 144 g 229 t 1 others
BASE COUNT	206 a 135 c 144 g 229 t 1 others
ALIGNMENT SCORES:	9,81e-16 Length: 715
Score: No:	236.00 Matches: 74
Percent Similarity:	45.06% Conservative: 40
Percent Local Similarity:	29.25% Mismatches: 69

FEATURES	source
REFERENCE	
TITLE	
JOURNAL	
COMMENT	
REFERENCE	
AUTHORS	
ORGANISM	
KEYWORDS	
ACCESSION	
VERSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	
LOCUS	
DEFINITION	
VERSION	
KEYWORDS	
ACCESSION	
DEFINITION	
LOCUS	
RESULT_18	
B0870484	


```

Db      482 CAAAGGAAAGTGGTAGTGGTTTATGATGAGCTGTATGAGGAGTCACCGCTAT 541
Qy      209 ARGAGPHEGINPROALALEU1LEASNAPVALPHEGINVALLEUGLUSERAPPROGU 228
Db      542 GTGCATTTTCCGCCGAAGTGAATTTCTTCCTCAGTTCGAGTTCAGCTACGATACCC--- 598
Qy      229 LysPheProThrGluLeuLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLys 248
Db      599 ---TTGCCAGAGTATCATGATCATTCCTAAAGAAATGGGACCATGACACGACTAG 655
Qy      249 CysThrLeuProTyrglyTlleLysLeuLeuMetGlyAsp-ValLys---GlyGlySerGI 267
Db      656 TTCTCTTAAACCACTTTGATGACTCTTGTACTGAGGAACCTGGAACCTGGTGTGATCATC 715
Qy      267 uThrPro 269
Db      716 CACACCC 722

RESULT 20
LOCUS   CD573451 515 bp mRNA linear EST 12-JUN-2003
DEFINITION 3529_1.127.1.E03_Y_1 3529 - 2 mm ear tissue from Schmidt and Hake
ACCESSION CD573451
VERSION   CD573451.1 GI:31664518
KEYWORDS  EST
SOURCE    Zea mays
ORGANISM  Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoidae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 515)
AUTHORS   Walbot V.
TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
          University
JOURNAL   Unpublished
COMMENT   Contact: Walbot V.
          Department of Biological Sciences
          Stanford University
          855 California Ave, Palo Alto, CA 94304, USA
          Tel: 650 723 2227
          Fax: 650 725 8221
          Email: walbot@stanford.edu
          Plate: 3529.1.127.1 row: B column: 03.

FEATURES
    source          location/Qualifiers
    1..515
    /organism="Zea mays"
    /mol_type="mRNA"
    /cultivar="B73"
    /db_xref="taxon:4577"
    /tissue_type="ear"
    /dev_stage="2 mm"
    /lab_host="E. coli XL0R"
    /lab_lib="3529 - 2 mm ear tissue from Schmidt and Hake
    labs"
    /note="Organ: ear; Vector: PAD-GAL4-2.1; Site 1: EORI;
    Site 2: XhoI; RNA isolated by Hake lab. 1 million pfu
    amplified. Ampicillin is the selection marker."

BASE COUNT 140 a 110 c 125 g 140 t
ORIGIN
Alignment Scores:
Pred. No.: 2,19e-14 Length: 515
Score: 222.00 Matches: 54
Percent Similarity: 52.98% Conservative: 35
Best Local Similarity: 32.14% Mismatches: 63
Query Match: 9.96% Indels: 16
Gaps: 3
US-10-069-062-7 (1-432) x CD573451 (1-515)
Qy 154 ValSerValAlaAlaThrSerLeuSerHisPheIle-Pro----- 167

```

```

Db      2 ATTTCTATTTCTTATAGAAAGACAGATGACACCGTGTCTCTTACTCCGAGACA 61
Qy      168 -----AenValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnI 183
Db      62 ATCTTGGCGACATGACATGACATGACAGAACTTGACTGTGTCACGCTATGGCCAAAG 121
Qy      183 eAlaHisCysTyraIaGlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTy 203
Db      122 TGACATTGTTTACACAGAGAAATTTGGACAGTGTTTATGATTTAGTGCTGTCTTA 181
Qy      203 rGlyLeuIleValIlyrArgarPheGlnProAlaLeuIleAsnAspValPheGlnVal-L 223
Db      182 TGGAGATCAACGCTATGATGATGATTTCTCCAGAAATACTCTCCGCGCTCAGGCTATAGG 241
Qy      223 euGIuserAspProGluLysPheProThrGluLeuLysLysLeuIleGluLysSerentpg 243
Db      242 TGGGCACT-----GTTCTGCCAGATGTATGATGATGATGATGATGATGATGATGATG 292
Qy      243 IuGIuLysHleGluArgCysThrLeuProTyrglyTlleLysLeuLeuMetGlyAspVal- 262
Db      293 ATCATGAGAAATAAACAGTTTCATTACCTCCTCGATGACCCCTTCTCTGGGGAACCTG 352
Qy      263 --LysGlyGlySerGluThrProLysLeuValSerArgValLeuGlnTrpLysLysGlu 282
Db      353 GGACCGGAGATCATCTATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 412
Qy      282 ySPProGluLysSerValValIlyrAspGlnLeuAsnSerAlaAsnLeuGlnInhemetL 302
Db      413 ACCCTGAGAAATCCAGAGATACATGAGTAAATCGCCATTGCCAATTCAACGCTGAGA 472
Qy      302 ySGuLeuArgGluMetArg 308
Db      473 ACCAATGAGATCTTAAAA 492

RESULT 21
LOCUS   B0869724 705 bp mRNA linear EST 15-AUG-2002
DEFINITION OGD7A12_Y9.AB1 OG ABCDI lettuce salinas Lactuca sativa cDNA clone
ACCESSION B0869724
VERSION   B0869724.1 GI:22255522
KEYWORDS  EST
SOURCE    Lactuca sativa
ORGANISM  Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; Campanulids; Asterales; Asteraceae; Cichorioideae;
Cichorieae; Lactuca.
REFERENCE 1 (bases 1 to 705)
AUTHORS   Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
          Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison
          P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, T., Lai, Z.,
          Church, S., Jackson, L. and Bradford, K.
          Lettuce and Sunflower ESTs from the Compositae Genome Project
          http://compgenomics.ucdavis.edu/
          Unpublished
JOURNAL   Unpublished
COMMENT   Contact: Alexander Kozik [R.W.Michelmore]
          Department of Vegetable Crops, R.W.Michelmore Lab
          University of California at Davis
          Asmndson Hall, UCD, Davis, CA 95616, USA
          Tel: 1-(530)-742-1742
          Fax: 1-(530)-752-9659
          Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
          belongs to contig OG_Ca_contig96356, see http://ccgdb.ucdavis.edu/
          for details.
          Plate: OGD7 row: A column: 12.

FEATURES
    source          location/Qualifiers
    1..705
    /organism="Lactuca sativa"
    /mol_type="mRNA"
    /cultivar="Salinas"
    /db_xref="taxon:4236"

```


www.reagen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 421.
Location/Qualifiers

FEATURES

source

1. .584
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl057-2963"
/tissue_type="Regenerating cotyledons, 2 week old seedling"
/lab_host="DH10B"
/clone_lib="Gm-cl057"
/note="Vector: pBluescript II SK⁺; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from mRNA isolated from degenerating cotyledons of 2 week old seedlings from p1468916. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

BASE COUNT 180 a 114 c 143 g 147 t
ORIGIN

Alignment Scores:

Pred. No.: 3.52e-13 Length: 584
Score: 212.00 Matches: 59
Percent Similarity: 48.74% Conservative: 38
Best Local Similarity: 29.65% Mismatches: 68
Query Match: 9.51% Indels: 34
Dbs: 12 Gaps: 7

US-10-069-062-7 (1-432) x BMS20154 (1-584)

QY 180 ValAlaGlnIleAlaHisCysTyrAlaGlnIleGlySerGlyPheAspValAla 199
Db 6 ATAGCTCAACTGACACACTGATGTCACAGGGGAAAGTTGGAGGTTATGTCGGC 65
QY 200 ThrAlaIleTyrGlyLeuIleValTyrArgArgPheGlnPheAlaLeuAsnAspVal 219
Db 66 TCAGCTGTGTATGGCAGTACGAGGCTATGCGGTTTTCACCGAGATGATTTCTTCCACT 125
QY 220 -----PheGlnValLeuGlnSerAspProGlnIlePheProThrGlnLeuIys 235
Db 126 AAGCTTGACATATAGCAGTGCCTTATCAGAT-----GGTATCACT 167
QY 236 LysLeuIleGlnSerAsnTrrGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 255
Db 168 GAAATTTATTAAGAAATTTGGACATGACACGACTGCTCTTACCACTTGTATG 227
QY 256 LysLeuMetGlyAspVal---LysGlyGlySerGlnIleProLysLeuValSerArg 274
Db 228 TCTCTTTTACTAGGAGAACCTGAGTGTGATATCAGCCATTCATGTTGTGTCT 287
QY 275 ValLeuGlnTrrLysGlyGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 294
Db 288 GTTAAATAAATGCAAAAGTCTGACCTCGAATCCCTGACACATGACAAAGATTGCA 347
QY 295 SerAlaAsn-----LeuGlnPheMetLysGlnLeu---ArgGlnMet 307
Db 348 GAGGCAAAATTCAGCATTTGAAATTCGAAATTCGAAATTCGAAATTCGAAATTCG 407
QY 308 ArgGlnIleTyrAspSer-----AspProGlnIleThrTyrIle 319
Db 408 TCGATGCAATTAATCTGTGATGACAGCTGACAGATCTCAGACACAGACAGATGAGAT 467
QY 320 LysGlnLeuAspHisSerValGlnProLeuThrVal----- 331
Db 468 GAACAAGCTTCTGAACCTTAACAAGAGACGATTATTAAGACACTGCTTGACAAAGA 527

QY 332 AlaIleLysAsnIleArgLysGlyLeuGlnAlaLeuThrGlnIleSerGlnValPro 350
Db 528 GCCATCTGGGAGTTATGATCATATATGCGCTATGGCGCAGCTCATGTCTTCA 584

RESULT 26

B0861999

LOCUS

DEFINITION

OGC1f02.yg.abl OG_ABCDI lettuce salinas

ACCESSION

B0861999

VERSION

B0861999.1 GI:22247464

KEYWORDS

EST.

SOURCE

ORGANISM

Lactuca sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

asterids; campanulids; Asterales; Asteraceae; Chloroidae;

Chloridae; Lactuca.

1 (bases 1 to 664)

Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,

Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison,

P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, T., Lai, Z.,

Church, S., Jackson, L. and Bradford, K.

Lettuce and Sunflower ESTs from the Compositae Genome Project

http://compositae.ucdavis.edu/

Unpublished

CONTACT: Alexander Kozik (R.W. Michelmore)

Department of Vegetable Crops, R.W. Michelmore Lab

University of California at Davis (UCD)

Asmundo Hall, UCD, Davis, CA 95616, USA

Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659

Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]

Belongs to contig Qc_Ca_contig196356, see http://cgdb.ucdavis.edu/

for details.

Plate: OGCI row: f column: 02.

FEATURES

source

1. .664

/organism="Lactuca sativa"

/mol_type="mRNA"

/cultivar="Salinas"

/db_xref="taxon:4236"

/clone="OGC1f02"

/lab_host="E.coli"

/clone_lib="OG_ABCDI lettuce salinas"

/note="Vector: pKODNASTAB; The library was constructed

from 10 different sources of RNA from a single genotype.

Separate cDNAs were generated using primers that

incorporated unique 5' and 3' tags to distinguish each

source of RNA. cDNAs were then pooled, size-fractionated,

directionally cloned into a custom medium-copy vector and

transformations made with four size classes to minimize

size bias. Details of each source of RNA and library

construction can be obtained at http://cgdb.ucdavis.edu/

TAG_Lib=OG_ABCDI lettuce salinas

TAG_Tissue=chemical induction

TAG_SEQ=TCRAGCGG"

BASE COUNT 193 a 125 c 136 g 209 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 4.28e-13 Length: 664
Score: 212.00 Matches: 68
Percent Similarity: 44.87% Conservative: 37
Best Local Similarity: 29.06% Mismatches: 59
Query Match: 9.51% Indels: 70
Dbs: 13 Gaps: 11

US-10-069-062-7 (1-432) x B0861999 (1-664)

QY 17 TyrLeuValLeuGlnIleProIleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHis 36
Db 17 TyrLeuValLeuGlnIleProIleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHis 36

Db 28 TATGCAATTGTAAGCCATTATGATGAA----- 57
 Qy 37 AlAValleIethrProlyserGlyThrserleuylsGlySerArglyleyleserPro 56
 Db 58 -----CTAACCCCAAGAGTTCCTTGGTCATGAGACAGATTGAAGTTAACTTCTCT 111
 Qy 57 GlnphealAaenglyglutrpilutyrhisleSerSerAsn-----ThrglyuylPro 74
 Db 112 CAATGGGCGCAG--GAAGCTATCTACAAATATGCTAAAGCATTTACTCTCAATGC 168
 Qy 75 ArggluValGlnserArglyleasnProphleuGlnuAlaThrllePhelelleValuAla 94
 Db 169 ACTTCACAGTATCAAG--AACCTTTGTGAAATGACAGTCAATATGCTGTACCA 225
 Qy 95 TyrileglnProthr-----Glu 100
 Db 226 GCAGCATATGCAACCTTTGCTATATGACAGAAGATGACCTTCAAAAACGCTTGAA 285
 Qy 101 AlAphaepleuGlnlelle-----lleYrSerAspProglyTyrhisSer 116
 Db 286 GGTATTGATATTAACAATTTAGTGGCAATGACTTTATTC-----TATGCAAT 336
 Qy 117 GlnGluAspThrGlu----- 121
 Db 337 GAGATTGAAGCAGCGGCTTCCACTAACCCCGAATCACTGGCATCTTACCACTTT 396
 Qy 122 -----ThrlYrThrsSerAsnGlyGlnYrThrPheleuYrhisSerArgla 138
 Db 397 TCTTCAATTAATCTTGAATCTTTCAGTGAAGAACT-----TGT 435
 Qy 139 lleThrgluValGlnYrThrglyleuGlySerSerAlaGlyleuValSerValAla 158
 Db 436 AAGCTGAAGTTGCAAAATCGGTTGGATCTTTCAGCCATGACACACAGCTGTT 495
 Qy 159 ThrsleuLeuSerHisPhe-----lleProAsnValleleSerThr----- 172
 Db 496 GCTGCTTACTTAATTAATCTTGAAGTTGATTTATCTTCACTTCAAGTGGGCCCAATAT 555
 Qy 173 -----AsnlyAspIleleuHisAsnValAlaGlnlelleAlaHisCysTyrAlaGlnlys 190
 Db 556 CAAAGAAATCTTGATGTGTGATTTATAGCCCAATGGCACGTGATTTGTCAGGA 615
 Qy 191 LysilegylSerGlyPheAspValAlaThraAlaIleYrGly 204
 Db 616 AAGTGTGAGTGTGATTTGATGTTAGTTCAGCTGTTATGGA 657
 RESULT 27
 BP272204 877 bp mRNA linear EST 07-MAR-2001
 LOCUS GA_EB0014D16f Gossypium arboreum 7-10 dpa fiber library Gossypium
 DEFINITION
 accession BP272204
 version BP272204.1 GI:11203199
 keywords EST.
 ORGANISM
 Gossypium arboreum
 Gossypium arboreum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
 REFERENCE
 AUTHORS Wing R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
 D., Wood,T.C., Leslie,A. and Wilkins,T.A.
 TITLE An integrated analysis of the genetics, development, and evolution
 of the cotton fiber
 JOURNAL
 COMMENT Unpublished
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: TAAATACGATCATATAGGG

FEATURES
 source
 1. 877
 /organism="Gossypium arboreum"
 /mol_type="mRNA"
 /strain="AKA"
 /cultiVar="8400"
 /db_xref="taxon:29729"
 /clone="GA_EB0014D16f"
 /issue_type="Fibers isolated from bolls harvested 7-10
 dpa"
 /lab_host="B. coli"
 /clone_id="Gossypium arboreum 7-10 dpa fiber library"
 /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 273 a 149 c 227 g 228 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.62e-12 Length: 877
 Score: 208.50 Matches: 63
 Percent Similarity: 46.25% Conservative: 48
 Best Local Similarity: 26.25% Mismatches: 94
 Query Match: 9.35% Indels: 35
 DB: 10 Gaps: 8
 US-10-069-062-7 (1-432) x BP272204 (1-877)
 Qy 208 TyrArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValleuGlnSerAspPro 227
 Db 3 TATGTCGTTTTTACCA-----GAAGTCTTTCGCTGCTCAGCTGACAGTG 50
 Qy 228 GlnYrPheProthrGlu-----LeuYrlySerleuIleGlnSerAsnThrglyGlnlys 245
 Db 51 AAAGGAGTCCACTAGAGAAATGATTTGGAACCTTTAAATGAAATGGACCAATGAG 110
 Qy 246 HisGlnArgCysThrleuProtyrGlylleuSerleuMetGlyAspVal-----LysGly 264
 Db 111 AGGACTGAATTCCTTTGCCACCATTAATGACATTTTACTGGGAGAACCGGAACCGT 170
 Qy 265 GlySerGluThrProlySleuValSerArgValleuGlnTrpYrlyGlnYrProgly 284
 Db 171 GATATCATCAACACCATCAATGTAGTGTCTTAAAAATGGCAGAGGCTGATCTGAA 230
 Qy 285 GlnSerSerValValYrAspGlnleuAsnSerAlaAsnleuGlnPheMetlyGlnleu 304
 Db 231 AATCCCAAGAAACATGAGAAAGCTGCCGATCTAATTCAGAACTTGACACACACTC 290
 Qy 305 ArgGlnMetArg-----GlnYrlySerAspSer-----Glu 313
 Db 291 AACATGTTAAGAAATTAAGCCAAAGAACACTGGATGCTTAATGCGTATGAAAAAC 350
 Qy 314 -----AspProGlnThrTyrIlelySerGlnleuAspHisSerVal-----Glu 327
 Db 351 TGTAGTAGCTTAAACACGACGAAGATGATGAGGAGTGAAGTGAAGTGAAGAGAA 410
 Qy 328 ProleuThrValAlaIlelys-----AsnleuGlyrSglyleuGln 341
 Db 411 GTTGTCAAATATTTGCTAAAGGAGGAGGAGGATTAAGTTCAGATCAAAATTCATCGCA 470
 Qy 342 AlaleuThrGlnYrSerGlnValProIleGlnProAspValGlnThrglnleuLeuAsp 361
 Db 471 ACATGGAGGAGGACAGATGTGTCATAGAACCCGAATCAACAGACTTAACTTTGGAT 530
 Qy 362 ArgCysGlnGlnleuProglyCysValGlyValValProGlyAlaGlyIYrAsp 381
 Db 531 GCTACTATGATATGAGAGAGTTTGTGGCCGAGTCTCTGACGATGTGATTTGAT 590
 Qy 382 AlaleuValleuValleuGlnAsnGlnValGlyAsnleuGlnYrThleuGln 401
 Db 591 GCAGCTTTTCTGTACCTTGGGTGAT-----TCTAGCAAGAACGTGACA 635
 Qy 402 AsnProAspYrPheHisAsnValYrTrpValAspleuGlnGlnThrglnGlyVal 421

Db		636 AATGAGTGAAGTCGCACATCTTCTGGCCTTGTTAGTAGGAAAGACCCTCAACTGT	695
RESULT 28			
B0049350			
LOCUS	B0049350	600 bp mRNA linear EST 26-AUG-2002	
DEFINITION	1111005H03.y1 1111 - Unigene III from Maize Genome Project Zea mays		
ACCESSION	B0049350	cDNA, mRNA sequence.	
VERSION	B0049350.1	GI:22489427	
KEYWORDS	EST.		
SOURCE	Zea mays		
ORGANISM	Zea mays		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD		
	clade; Panicoidae; Andropogoneae; Zea.		
REFERENCE	1 (bases 1 to 600)		
AUTHORS	Walbot,V.		
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford University		
JOURNAL	Unpublished		
COMMENT	Contact: Walbot V Department of Biological Sciences Stanford University 855 California Ave, Palo Alto, CA 94304, USA Tel: 650 723 2227 Fax: 650 725 8221 Email: walbot@stanford.edu Plate: 1111005 row: H column: 03. Location/Qualifiers 1..600 /organism="Zea mays" /mol_type="mRNA" /db_xref="dbEST:952012C01.y1" /db_xref="taxon:4577" /clone_lib="1111 - Unigene III from Maize Genome Project" /note="This library represents the unique genes found in the third round of EST sequencing at Stanford University for the maize genome project. Sequences are present from library 952. Contigs were assembled using ZmBAssembler and 2 representatives from each contig were selected for the unigene set. All singlets were also selected."		
FEATURES	source		
	153 a 132 c 131 g 184 t		
BASE COUNT			
ORIGIN			
Alignment Scores:			
Pred. No.:	1.17e-12	Length:	600
Score:	207.50	Matches:	55
Percent Similarity:	49.11%	Conservative:	28
Best Local Similarity:	32.54%	Mismatch:	51
Query Match:	9.30%	Indels:	35
DB:	13	Gaps:	6
US-10-069-062-7 (1-432) x B0049350 (1-600)			
Oy	109 TyrsGrAPProGLYTYRHISserGlnGUaPThrGluThRLysThSerSerangly	128	
Db	148 TTCCTTCAATTACATTAACTGACAAGTGTGCTAATGGAACATGACT-----CGA	198	
Oy	129 GluLYSThrPhelLeuTYRHisSerArgAlaIleThrGluValGluLYSThrGlyLeugly	148	
Db	199 GAGAAA-----TGTAAACCTGAAGTGGCCAAACTGGACTGGG	237	
Oy	149 SerSerAlaGlyLeuValSerValAlaThrSerLeuLeuSerHisPheIle-----	166	
Db	238 TCATAGAGTGCATATCACCATCATGATTGTGAGCGCTTTCTT---CACATCTTGATGTC	294	
Oy	167 -----ProGnValILIsEThraNlysAp	175	
Db	295 GTTAAGCTCTCATGTCCAGACAACTCTTCTGTGATAATACAACTAGACGAACCTTGAC	354	
Oy	176 ILeuNHISerValAlaGlnIleAlaHisCytyrAlaGlnLysLytleGlysergly	195	

DB	Accession	Version	Keywords	Source	Organism	Reference	Authors	Title	Journal	Medline	PubMed	Comment
DB	355	TTGGTTCACCTTATTTGCCCAAGGACATTTGATATACACAGGGGAAATTGGCAGTGT	414									
OY	196	PheapsyValalattirAlaIleTyrTgLYleuIleValTyrTgATgATgPheGlnProAlaLeu	215									
DB	415	TTTTATGTTAGTGGCTGGCTCTATGAGGAGTCAACGCTATGATGTAAGTTTCTCCA	468									
OY	216	IleasnAspValPheGlnValIleuGluSerAspProGluLys	231									
DB	469	-----GAAATACCTTCCTCTCTCAGGCTACAGGTGACCTTCTCCCA	513									
OY	232	ThrGluLeuValysLeuIleGluSerAsnTyrPheGluLysPheGluGlyCysThrLeu	251									
DB	514	GATGATGATACAGATATGTTTACACAAAGGTGGATCATGAGATAAACAAGTTCTATTA	573									
OY	252	ProTyrGlyIleLeuLysLeuMetGly	260									
DB	574	CTCTCTGATGACCTTCTCTTGGG	600									
RESULT 29												
LOCUS	AO911417/c											
DEFINITION	AO911417	594 bp	DNA	linear	GSS 25-MAY-2001							
ACCESSION	AO911417.1	GI:6507933										
VERSION	AO911417.1	GI:6507933										
KEYWORDS	GSS.											
SOURCE	Leishmania major											
ORGANISM	Leishmania major											
REFERENCE	1 (bases 1 to 594)											
AUTHORS	Akopyants,N.S., Clifton,S.W., Martin,J., Pape,D., Wylie,T., Li,L., Kassing,J.C., Roos,D.S. and Beverley,S.M.											
TITLE	A survey of the Leishmania major Friedlin strain VI genome by shotgun sequencing: a resource for DNA microarrays and expression profiling											
JOURNAL	Mol. Biochem. Parasitol.	113 (2),	337-340	(2001)								
MEDLINE	11295190											
PUBMED	11295190											
COMMENT	Contact: Akopyants, NS / Beverley, SM Washu Leishmania Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu Library construction: Natalia S. Akopyants, Ph.D. DNA sequencing by: Washington University Genome Sequencing Center If using this information please cite: N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major Friedlin strain VI genome by shotgun sequencing' and the Washington University Genome Sequencing Center For information on obtaining clone material please contact: Natalia S. Akopyants Ph.D. (natalia@wustl.wustl.edu) and/or Stephen M. Beverley Ph.D. (beverley@borcim.wustl.edu) Seq primer: -40UP from G1BCO Class: Shotgun High quality sequence stop: 317. Location/Qualifiers 1. 594 /organism="Leishmania major" /mol type="genomic DNA" /strain="Friedlin strain VI" /db xref="taxon:5664" /clone="LMAJFV1_in04d05" /lab host="TOP10 (Invitrogen)" /clone_id="Leishmania major FV1 random genomic library" /note="Vector: pZero-2 (Invitrogen); Site 1: EcoRV; Genomic DNA was isolated from stationary phase cells. For this library, DNA was sheared to give a tight size distribution of 1-1.5kb fragments, blunt-ended with T4 DNA											

polymerase, dephosphorylated with Shrimp Alkaline
Phosphatase and ligated into pZero-2 vector's EcoRV site."

ALIGNMENT SCORES:

Pred. No.: 2,56-12 Length: 594
Score: 204.50 Matches: 62
Percent Similarity: 44.95% Conservative: 36
Best Local Similarity: 28.44% Mismatches: 80
Query Match: 9.17% Indels: 40
DB: 28 Gaps: 7

US-10-069-062-7 (1-432) x AG911417 (1-594)

QY 180 ValAlaGlnIleAlaHisCysTyrAlaGlnIleGlySerGlyPheAspValAla 199
DB 594 ATAGCCATGATGCTCTACAGGCTCAGACAGGCGCAATAGAGCGGCTTCAGCTGAC 535
QY 200 ThrAlaIleTyrGlyLeuIleValTyrArgArgPheGlnProAlaLeuIleAsnAspVal 219
DB 534 ACCGAGCTACAGGCAATGCGCTACCGCTGCTATCTGTGAGCGCGGTATCATGATG 475
QY 220 PheGlnValLeuGluSerAspProGluIlySphProThrGluLeuIleGlu 239
DB 474 ATGATGCGCAGCGCTCAGCCAGCTCGGTGAGTACAG-----ACATTTGAGC 424
QY 240 SerAsn-TyrGluGluIlyS-----HisGluArgCysThrLeuProTyrGly 254
DB 423 GCTTTGGTGGAGCTGAAATGTGTTGGTCCCGCTGAGTGTTCGCGCTTACGCGCTG 364
QY 254 YIleIlyLeuLeuIleGlyAspVal-----LysGlyGlySerGluThrProIlyLeuValse 273
DB 363 AGTAAAGCTGTGCTTGGAGAGCTGACCAAGGCGGCTCTGCACTCTGTGATGCTTGC 304
QY 273 TArgValLeuGlnIlyIlySlyGlyIlySphProGluIlySerValValTyrAspGlnIle 293
DB 303 AAAGATCATGCGCTGCGCAAGCTGTGCTGACACCCAGACCAACCTTTGGAAACAGCT 244
QY 293 uAsnSerAlaIleuGlnIlyPheMetIlyGluLeuArgGluMetArgIlySlyTyrAspSe 313
DB 243 GCGTGGCAACATAGAGCGGTACATGCGCTGTGCTGATGATGACGAGCGGCGAC 184
QY 313 TAspProGluIlyThrTyrIleIlySlyGluLeuAspHisSerValGluProLeuThrValAlaIle 333
DB 183 GAAGCGGAGCTGAC----- 168
QY 333 eLysAsnIleArgIlySlyLeuGlnAlaLeuThrGlnIlySerGluValProIle---Gly 352
DB 167 -----GCGGCTTCTATGAGCGCTCAGACAGGTGCGCTGCGCTGCGCTGTTCA 118
QY 352 uProAspValGln-----ThrGlnLeuLeuAspArgCysGly 364
DB 117 GCGACAGCGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 58
QY 364 nGluIlePro-----GlyCysValGlyIlyValValProGluIlyGly 378
DB 57 TGCATTCCTTGGGACATGAGTGTGAGCGGCGGAGTGAAGTGAAGTGAAGTGAAGTGA 6

RESULT 30
BM276808
LOCUS 612 bp, mRNA linear EST 28-JAN-2002
DEFINITION 95201C01.y1 952 - BMS tissue from Walbot Lab (reduced RNA) Zea
ACCESSION BM276808
VERSION BM276808.1 GI:17970040
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 612)

AUTHORS Walbot, V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL University
COMMENT Unpublished
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 952012 Row: C Column: 01.
Location/Qualifiers

FEATURES

source
1. .612
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="BMS (Black Mexican Sweet)"
/db_xref="taxon:4577"
/tissue_type="suspension culture"
/dev_stage="mixed logarithmic and stationary growth
phases"
/lab_host="DH10B"
/clone_id="952 - BMS tissue from Walbot Lab (reduced RNA
)"

/note="Vector: pUC19; Site 1: EcoRI; Site 2: EcoRI; The
library was prepared by George Rudenko using poly (A)
selected RNA and Universal Riboclone cDNA Synthesis System
(Promega). cDNA was synthesized using both random and
oligo(dT) primers in separate reactions and equipped with
EcoRI adaptors. Library was size-fractionated on agarose
gels (for insert size >400bp) and non-directionally cloned
into EcoRI-digested pUC19 vector. Blue/white selection on
carbenicillin-containing plates was used to recover
positive clones."

BASE COUNT 156 a 136 c 134 g 185 t 1 others
ORIGIN

ALIGNMENT SCORES:

Pred. No.: 2,626-12 Length: 612
Score: 204.50 Matches: 55
Percent Similarity: 48.52% Conservative: 27
Best Local Similarity: 32.54% Mismatches: 52
Query Match: 9.17% Indels: 35
DB: 12 Gaps: 6

US-10-069-062-7 (1-432) x BM276808 (1-612)

QY 109 TyrSerAspProGluTyrHisSerGlnIlyAspThrGluThrIlyThrSerSerAsnGly 128
DB 158 TTCTCTTCATTTACATTTAATCTAGAGGTGCTAATGCAATGACT-----GGA 208
QY 129 GluIlyThrPheLeuTyrHisSerArgAlaIleThrGluValGluIlyThrGlyLeuGly 148
DB 209 GAGAAA-----TGTAACCTGTAAGTACCAAACTGACCTTGGG 247
QY 149 SerSerAlaGlyLeuValSerValAlaIleThrSerIleuSerHisPheIle----- 166
DB 248 TCATTCAGCTCCATGACACATCATGATGTTGTCAGCTTCTT---CATATCTTGGTTC 304
QY 167 -----ProAsnValIleSerThrAsnIlyAsp 175
DB 305 GTTAGCCTTCATGTCAGACCAATCTTGTGATGATATCAATGACAGCAACTTGAC 364
QY 176 IleLeuHisAsnValAlaGlnIleAlaHisCysTyrAlaGlnIlySlyIleGlySerGly 195
DB 365 TTGGTTCATCTATGCGCAAAAGTGCACATTTATGACAGGCGAAATTTGCGAGTGT 424
QY 196 PheAspValAlaThrAlaIleTyrGlyLeuIleValTyrArgArgPheGlnProAlaLeu 215
DB 425 TTGATGTTTATGCTGCTGCTATGAGAGTCAACGCTATATTAAGTTTCTCCA----- 478
QY 216 IleAsnAspValPheGlnValLeuGluSerAspProGluIlyS-----PhePro 231

Db	479	-----GAATACTCTCCCTGCTCAGGCTACAGGNGCATTCTCCCA	523
Oy	232	Thrltuleuyls/lsleuilegluservantrrpulgulysrhlsgluarCyThrlieu	251
Db	524	GATGTAGTATCAGATTTGTTATTCACCAAGAGGTCATGAGAAATTAACGCTCATTA	583
Oy	252	ProTyrGlyIleLysLeuLeuMetGly	260
Db	584	CTCTCTGATGACCTCTTCTCTGGG	610
RESULT 31			
LOCUS	AO911416/c		
DEFINITION	LMJFV1_in04d04_x1 Leishmania major FV1 random genomic library		
VERSION	AO911416		
KEYWORDS	SW:ER08_YEAST P24521 PHOSPHOMEVALONATE KINASE ;, genomic survey		
SOURCE	AO911416.1 GI:6507932		
ORGANISM	GSS.		
REFERENCE	Leishmania major		
AUTHORS	Leishmania major		
TITLE	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.		
JOURNAL	1 (bases 1 to 608)		
MEDLINE	Akopyants,N.S., Clifton,S.W., Martin,J., Pape,D., Wylie,T., Li,L.,		
PUBMED	Kissinger,J.C., Roos,D.S. and Beverley,S.M.		
COMMENT	A survey of the Leishmania major Friedlin strain V1 genome by shotgun sequencing: a resource for DNA microarrays and expression profiling		
	Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)		
	21192569		
	11295190		
	Contact: Akopyants, NS / Beverley, SM		
	Washu Leishmania Project		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	Library construction: Natalia S. Akopyants, Ph.D.		
	DNA Sequencing by: Washington University Genome Sequencing Center		
	If using this information please cite:		
	N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major Friedlin strain V1 genome by shotgun sequencing' and the Washington University Genome Sequencing Center For information on obtaining clone material please contact: Natalia S. Akopyants Ph.D. (natalia@orcin.wustl.edu) and/or Stephen M. Beverley Ph.D. (beverley@orcin.wustl.edu)		
	Seq primer: -40UP from Gibco		
	Class: Shotgun		
FEATURES	High quality sequence stopt: 302.		
source	Location/Qualifiers		
	1. 608		
	/organism="Leishmania major"		
	/mol_type="genomic DNA"		
	/strain="Friedlin strain V1"		
	/db_xref="taxon:5664"		
	/clone="LMJFV1_in04d04"		
	/lab_host="TOP10 (Invitrogen)"		
	/clone_lib="Leishmania major FV1 random genomic library"		
	/note="Vector: pZero-2 (Invitrogen) ; Site 1: EcoRV;		
	Genomic DNA was isolated from stationary phase cells. For		
	this library, DNA was sheared to give a tight size		
	distribution of 1-1.5kb fragments, blunt-ended with T4 DNA		
	polymerase, dephosphorylated with Shrimp Alkaline		
	Phosphatase and ligated into pZero-2 vector's EcoRV site."		
BASE COUNT	106 a 199 c 188 g 112 t	3 others	
ALIGNMENT SCORES:	1.22e-11	Length: 608	
Pred. No.:	198.50	Matches: 49	
Score:			

Percent Similarity:	52.29%	Conservative:	31
Best Local Similarity:	32.03%	Mismatches:	58
Query Match:	8.90%	Indels:	16
DB:	28	Gaps:	3
US-10-069-062-7 (1-432) x A0911416 (1-608)			
Qy 175 AsplLeuHsAsnValAlaGlnIleAlaHsCysTrpAlaGlnLysLysIleGlySer	194		
Db 605 GAGTACGTCATCATCTGATGCCCAATGCTCTCAAGCTCACACAGACAAATCGGACG	546		
Qy 195 GlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyrArgArgPheGlnProAla	214		
Db 545 GCGTTGACCGTATACACCGCGCTTACAGCAATGCGCCCTCCGCGCTTTCGTGACG	486		
Qy 215 LeuIleAsnAspValPheGlnValLeuGlnSerAspProGlnLysPheProThrGln	233		
Db 485 CGCGTATCGATGATGATGATGAGCGACGCGCTACGCCA-CCGAGCTCGGTCGAGTGAACACA	427		
Qy 234 -----LeuLysLysLeuIleGlnSerAspTrpGlnGlnLysPhe	246		
Db 426 CTAAAGCCGGTGGGACAAATGACCAAGGCTGTGGGTCCGCCCAAGTGG-----	379		
Qy 247 GlnArgCysThrLeuProTyrGlyIleLysLeuLysMetGlyAspVal--LysGlyGly	265		
Db 378 -----TTCCGCGCTTACCGCTCGAGTAACTGTCTCGGACACGTGCACCAAGCGGC	325		
Qy 266 SerGlnThrProLysLeuValSerArgValLeuGlnTrpLysLysGlnLysProGlnGln	285		
Db 324 TCTGCACTCTCGGTATAGTTGCAAAATCATGAGCGTGGCGCAAGTCTGTGCTGACACC	265		
Qy 286 SerSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGlnLeuArg	305		
Db 264 CCAGACAACTTTGGGAAACAGCTGCGTGGCAACATGAGCGTACAGCGCGCTTGGCT	205		
Qy 306 GlnMetArgGlnLysTyrAspSerAspProGlnThrTyr	318		
Db 204 CGCATGATTCACAGAGCGGACGACGACCGGACGCTTAC	166		
RESULT 32			
LOCUS B1972859	549 bp	mRNA	linear
DEFINITION B1972859 y1 Gm-c1065 Glycine max cDNA clone GENOME SYSTEMS CLONE			
ID: Gm-c1065-7669 5' similar to SW:BR03_YEAST P24521			
PHOSPHOMEVOLONATE KINASE ; mRNA sequence.			
ACCESSION B1972859			
VERSION B1972859.1	GI:16347264		
KEYWORDS EST.			
SOURCE Glycine max (soybean)			
ORGANISM Glycine max			
REFERENCE Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
AUTHORS SpERMtophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids			
1 (bases 1 to 549)			
Shoemaker, R., Keim, P., Vodkin, L., Erpelidg, J., Coryell, V., Khanna			
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,			
Wyle, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers			
Y., Person, B., Swaller, T., Gibson, M., Pape, D., Harvey, N., Schunk			
R., Ritters, E., Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McMan			
R., Waterson, R. and Wilson, R.			
Public Soybean EST Project			
Unpublished			
Contact: Shoemaker R/Public Soybean EST Project			
Washington University School of Medicine			
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA			
Tel: 314 286 1800			
Fax: 314 286 1810			
Email: est@wustl.edu			
This clone is available through: ResGen, Invitrogen Corp. 2130			
South Memorial Parkway Huntsville, AL 35801 For further information			
call: (800)-533-4363 or contact via email: cc@resgen.com			
TITLE			
JOURNAL			
COMMENT			


```

ACCESSION      Gm-cl065-2295.5', similar to SW:ERGB_YEAST P24521 PHOSPHOMEVALONATE
KEYWORD        KINASE ;, mRNA sequence.
VERSION        BF070746.1
KEYWORDS       BF070746.1 GI:10848203
SOURCE         EST.
ORGANISM       Glycine max (soybean)
               Glycine max
REFERENCE      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS        Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
               ; eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
               Glycine.
TITLE          1 (bases 1 to 405)
JOURNAL        Shoemaker,R., Kelm,P., Vodkin,L., Expediting,J., Corvelli,V., Khanna
COMMENT        A., Bolla,B., Marr,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
               Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
               ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurr
               ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
               ,R., Waterston,R. and Wilson,R.
               Public Soybean EST Project
               Unpublished
               Contact: Shoemaker R/Public Soybean EST Project
               Public Soybean EST Project
               Washington University School of Medicine
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: est@watson.wustl.edu
               This clone is available through: ResGen, Invitrogen Corp. 2130
               South Memorial Parkway Huntville, AL 35801 For further information
               call: (800)-533-4363 or contact via email: ccu@resgen.com
               Insert Length: 1534 Std Error: 0.00
               High quality sequence stop: 301.
               Location/Qualifiers
               1..405
               /organism="Glycine max"
               /mol_type="mRNA"
               /db_xref="taxon:3847"
               /clone="GENOME SYSTEMS CLONE ID: Gm-cl065-2295"
               /tissue_type="germinating shoots"
               /lab_host="DH10B"
               /clone_idb="Gm-cl065"
               /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
               XhoI; The cDNA library was constructed from mRNA isolated
               germinating shoots of the cultivar Williams. The seeds
               were allowed to germinate for 24 hours prior to being
               cold stressed for 2 days at 4C. Complementary DNA was
               synthesized from mRNA using a primer consisting of a
               poly(dT) sequence with a XhoI restriction site. EcoRI
               adapters were ligated to the blunt-ended cDNA fragments
               followed by XhoI digestion. The cDNA fragments were
               directionally cloned into the EcoRI-XhoI restriction
               site of the pBluescript vector. The ligated cDNA fragments were
               transformed into DH10B host cells (GibcoBRL). This library
               was constructed in the laboratory of Dr. Randy
               Shoemaker."
BASE COUNT     96 a      87 c      103 g      119 t
ORIGIN
Alignment Scores:
Pred. No.:      6,786-10      Length:      405
Score:          180.50      Matches:      40
Best Local Similarity: 62.92%      Conservative: 16
Percent Similarity: 44.94%      Mismatches:  20
Query Match:    8.09%      Indels:      13
DB:             10      Gaps:      3
US-10-069-062-7 (1-432) x BF070746 (1-405)
QY      141 GluValGluValThGlyLeuGlySerSerAlaGlyLeuValSerValAlaAlaThrSer 160
Db      103 GAAGGCGCTAAACACGTTTGCGCTCTTCGTGACGAGTACACATGCGCTGTAGTTCGCT 162
QY      161 LeuLeuSerThrIlePheIleProIleVal---IleSerThrIleValys----- 174

```

QY	Db	163	CTACTT---	CATTACTGAGTATGTTAAAGCTTTCCTTGGAAGATCATCGGAAAG	219
QY	175	-----	-----	Asp1LeuHisAnValAlaGlnIleValHisCysTyrAla	188
Db	220	AAGATGTTGCAGATCTTATGATATGTCGATATAAATAGCTCAAACTGCACACTGATATGCA	279		
QY	189	GlnLysLysIleGlySerGlyPheAspAlaLysThrIleIleTyrGlyLysIleValTyr	208		
Db	280	CAGGGAGAAAGTGGCAGTGGGTTAATATGATTCAGCTCAGCTGTGTATGGCAGTCACTGCTAT	339		
QY	209	ArgArgPheGlnProAlaLeuIleAsn	217		
Db	340	GTGCGGTTTTCACCGGAAGTATTTCT	366		
RESULT 35					
LOCUS	B0110266	750 bp	mRNA	linear	EST 01-APR-2000
DEFINITION	VD0107C11 VD01 Verticillium dahliae cDNA, mRNA sequence.				
ACCESSION	B0110266				
VERSION	B0110266.1	GI:29426636			
KEYWORDS	EST.				
ORGANISM	Verticillium dahliae				
REFERENCE	1 (bases 1 to 750)				
AUTHORS	Neumann,M.J. and Dobinson,K.F.				
TITLE	Sequence tag analysis of gene expression during pathogenic growth and microsclerotia development in the vascular wilt pathogen Verticillium dahliae				
JOURNAL	Fungal Genet. Biol. 38 (1), 54-62 (2003)				
MEDLINE	22442000				
PUBMED	12553936				
COMMENT	Contact: Dobinson KF Agriculture & Agri-Food Canada 1391 Sandford St., London, Canada, N5V 4T3 Email: dobinsonk@em.agr.ca. Location/Qualifiers 1. 750				
FEATURES					
SOURCE	/organism="Verticillium dahliae" /mol_type="mRNA" /strain="Dvd-T5" /db_xref="taxon:27337" /lab_host="E. coli" /clone_1ib="VD01" /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Developing microsclerotia (DMS) cDNA library; constructed from axenic cultures grown on cellulose membranes overlaid onto a basal medium agar. Cells were harvested at 4 days post-inoculation. Single-pass sequencing was done using the T3 promoter primer: 5' ATTACCCCTCAATAAGGA 3'."				
BASE COUNT	135 a	278 c	199 g	134 t	4 others
ORIGIN					
Alignment Scores:					
Pred. No.:	1.75e-09	Length:	750		
Score:	180.50	Matches:	58		
Percent Similarity:	45.218	Conservative:	27		
Best Local Similarity:	30.854	Mismatches:	80		
Query Match:	8.09%	Indels:	23		
DB:	13	Gaps:	7		
US-10-069-062-7 (1-432) x B0110266 (1-750)					
QY	4	AlaPheSerAlaProGlyLysAlaPheLeuAlaGlyIleTyrLeuValLleGluProIle	23		
Db	124	GCCGTCGCCGCCCGGAGCAAGTCTCCCTCGGCGGCGGCTACCTGTCCTCGACGCGGC	183		
QY	24	TyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaVal-----IleThr	40		

```

Db      184 TACAGGCGCTTGTGTTGGGCTGAGCGCGGCATCAGCGTCTGCGCGGAGCTGACC 243
Qy      41 ProlysgIyThrSerLeuIysGluSerArgIleYrHisSerSerProGlnPheIAsn 60
Db      244 TCTGGCGTTCGAGCGGCTGACGTGAGCGAGGTTCGTGAGAGCGCCGACGTTGAGGCT 303
Qy      61 GlyIuIyPrgIuYr-----HisIleSerSerAnthrIuIyPrgArgIuValGln 78
Db      304 GCCGGGTGGCGGTATGAGTTCGCGCTGGGAGGGGGGCGTGCAGGTATACGAGGTGACG 363
Qy      79 SerArgIle-----AsnProPheLeuGluIuIaThrIlePheIleValIeuAla 94
Db      364 GTTGGCGCTTCATCACAACCAACCCCTTTGTGAGAAACACATCCTACCGCCCTCACCC 423
Qy      95 TyrIle-----GlnProThrGluIuIaPheAspLeu-----GluIleIle 107
Db      424 TACATTCACGAGATTGTGTCGCCAACAACACTCTACAGCCTCGCTCGCTCGTCTCACCC 483
Qy      108 IleTyrSerAspProGlyIyThrHisSerGlnIuIaPrgIuIyThrIyThrSerSerAsn 127
Db      484 ATTGACGCCGACAACGACTACTACACGTTGGCCCGGAGCGCGTTACACAGACTGGCGCT 543
Qy      128 GlyIuIyThrPheLeuYrHisSerArgAlaIleThrGluValGluIyThrGlyLeu 147
Db      544 -----TACGCAACCCCTACACCTTCACCTCGCGCGGCGCACAGACGGCGCTCG 591
Qy      148 GlySerSerAlaGlyLeuValSerValIuIaIaThrSerLeuIeuSerHisPheIle--- 166
Db      592 GGTCTCTCCGCGCGCTGCTGACCTTCCTTCACCGCGCGCTGATGAGCAGCTATCTGACC 651
Qy      167 -----ProAsnValIleSerThr 172
Db      652 ACGTTCCTCTCTGCACTCAACC 675

RESULT 36
BQ869223      639 bp mRNA linear EST 14-AUG-2002
LOCUS      OGD5113.Y9.ab1 OG ABCDI lettuce salinas Lactuca sativa cDNA clone
DEFINITION
ACCESSION      BQ869223
VERSION      BQ869223.1 GI:22254980
KEYWORDS      EST.
SOURCE      Lactuca sativa
ORGANISM      Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asterales; Cichorioideae;
Cichorieae; Lactuca.
REFERENCE
AUTHORS      Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., Van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Lecture,S., Jackson,L. and Bradford,K.
TITLE      Lettuce and Sunflower ESTs from the Compositae Genome Project
COMMENT      http://compgenome.ucdavis.edu/
Unpublished
Contact: Alexander Kozik (R.W.Michelmore)
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Amundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
belongs to contig OG_Ca_Contig6356, see http://cgpdb.ucdavis.edu/
for details.
Plate: QGD5 row: 1 column: 13.

FEATURES
source
location/Qualifiers
1..639
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="OGD5113"

```

```

/lab host="E.coli"
/clone.lib="OG ABCDI lettuce salinas"
/notes="Vector: pBRCNDNA5flab: The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_LIB=OG ABCDI lettuce salinas
TAG_TISSUE=chemical induction
TAG_SEQ=GTAGCCGGG"

BASE COUNT      186 a 122 c 132 g 199 t
ORIGIN

Alignment Scores:
Pred. No.:      1,566-09      Length:      639
Score:          180.00      Matches:      64
Percent Similarity: 43.86%      Conservative: 36
Best Local Similarity: 28.07%      Mismatches: 57
Query Match:    8.07%      Indels:      71
DB:             13      Gaps:      11

US-10-069-062-7 (1-432) x BQ869223 (1-639)

Qy      17 TyrIeuValIeuGluProlIeYrAspAlaIyValIaIeuSerSerArgMetHis 36
Db      28 TATGCAATTGTAAAGCCAAATTATGATGAA-----
Qy      37 AlaValIleThrProlysgIyThrSerLeuIysGluSerArgIleYrHisSerSerPro 56
Db      58 -----CTAACGCCAACAAGCTGCTTGGCTATGACAGATTGAAAGTTACTTCTCT 111
Qy      57 GlnPheAlaAsnGlyIuIyPrgIuIyThrHisSerSerAsn-----ThrGluIyPrg 74
Db      112 CAATAGGCGAGA---GAAGCATCTACAATAATGCTAAAGCATTTTACTCTTCATTC 168
Qy      75 ArgGluValIeuSerArgIleAsnProPheLeuGluIuIaThrIlePheIleValIeuAla 94
Db      169 ACTTCAGTAATCAAG---AACCTTTGTGGAATATGCAATGCAATATGCTGTAGCA 225
Qy      95 TyrIleGlnProThr-----Glu 100
Db      226 GCAGCATATGCAACATTGGTATATGACAAGAGATCCTTCAAAACTGCTTGAAA 285
Qy      101 AlaPheAspLeuGluIleIle-----IleTyrSerAspProGlyIyThrHisSer 116
Db      286 GGTTCATATTAATTAATTTAGGTGCGAATGACTTTTATTC-----TATCGCAAT 336
Qy      117 GlnGluAspThrGlu----- 121
Db      337 GAGATTGAAACACAGTGGCTTCACCTAACCCGGAATCAGTGCATCTTACCACTTTT 396
Qy      122 -----ThrIyThrSerSerAnthrIuIyPrgIuIyThrPheLeuYrHisSerArgAla 138
Db      397 TCTTCAATTACTTTGAATTCTTCAAGTGAAGAAAGT-----TGT 435
Qy      139 IleThrGluValGluIyThrGlyLeuGlySerSerAlaGlyLeuValIeuValIaIa 158
Db      436 AAGCGTAATTCGAAAAAATGGGTGGATCTTTCACACCATGACACACAGAGTGTT 495
Qy      159 ThrSerLeuIeuSerHisPhe---IleProAsnValIleSerThr----- 172
Db      496 GCGCTTACTTATTAATCCTTGAGGTGTTAACTTCACTTCACTTGTGGGCGCCAAATAT 555
Qy      173 -----AsnIyAspIleIeuHisAsnValIaIaGlnIleIaIaIyCys-TyrIaIaGlnIy 190
Db      556 CAAGAAATCTTATGTGTGTCATATTATAGCCAAATGACACATGTATTTGCTCAGGA 615
Qy      190 sIyIleIleGlySerGlyPheAsp 197

```

DB	RESULT 37	616	AAAGTTGGAGTGGATTGAT 637
DB	LOCUS	BGS95511	617 bp mRNA linear EST 07-MAR-2003
DB	DEFINITION	EST494189 cSTS Solanum tuberosum cDNA clone cSTS10L11 5' sequence,	
DB	ACCESSION	BGS95511	
DB	VERSION	BGS95511.1	GI:13613651
DB	KEYWORDS	EST.	
DB	ORGANISM	Solanum tuberosum (potato)	
DB	SOURCE	Solanum tuberosum	
DB	REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Solanum.	
DB	AUTHORS	van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Chiemingo, A., Bongori, O., Buell, C. R., Romling, C., Tanksley, S. and Baker, B.	
DB	TITLE	Generations of ESTs from sprouting potato eyes	
DB	JOURNAL	Unpublished	
DB	COMMENT	Contact: Robin Buell The Institute for Genomic Research 9712 Medical Center Dr, Rockville, MD 20850, USA Email: potato-array@igr.org This clone can be obtained from the University of Arizona Genomics Institute. Orders can be made through URL: http://genome.arizona.edu/orders/ Seq primer: M13F-R.	
DB	FEATURES	Location/Qualifiers	
DB	source	1..617 /organism="Solanum tuberosum" /mol_type="mRNA" /cultivar="Kennebec" /db_xref="taxon:4113" /clone="cSTS10L11" /tissue_type="sprouting eyes from tubers" /dev_stage="12-14 weeks post harvest" /lab_host="SOLR" /clone_idb="cSTS" /note="vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Various sizes of sprouting eyes (2mm to 15mm) were taken from tubers. The tubers were incubated at 26C in the dark for 2-3 weeks prior to sprouting. The eyes were frozen in liquid nitrogen immediately upon removal from tubers."	
DB	BASE COUNT	162 a 137 c 144 g 174 t	
DB	ORIGIN		
DB	Alignment Scores:		
DB	Pred. No.:	1,68e-09	Length: 617
DB	Score:	179.50	Matches: 54
DB	Percent Similarity:	42.25%	Conservative: 25
DB	Best Local Similarity:	28.88%	Mismatches: 47
DB	Query Match:	8.05%	Indels: 61
DB	BB:	10	Gaps: 7
DB	US-10-069-062-7 (1-432) x BGS95511 (1-617)		
DB	QY	82	AsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrlleGlnProThr----- 99
DB	DB	36	AAACCCCTTGTAGAACATGACGTGAGATGCTGTTCGACGACGCCCAATGACACTTTCGAC 95
DB	QY	100	-----GluAlaPheAspLeuGluIleIle 107
DB	DB	96	AAAGTAGAAGACACATTCGACAGACTTCTTCGACGGGCTTGACATACATATCTTCG 155
DB	QY	108	-----IleTyrSerAspProGlyTyrHisSerGlnGluAsp----- 119
DB	DB	156	GGTTGCATGAGATTCTATTCG-----TATCGAATCAGATTGAAGCAGCTGGGCTC 206
DB	QY	120	-----ThrluThrIlys 123
DB	DB	207	CCTCTTAACCTTAAGTCCCTGGCTTCCTTCACACTTTTCTTCGATCATCCTTTATATGCA 266

QY	124	ThrserserAangnglygluylThrsPhelenuThrsiserrAgatallThrglVlaGl	143
Db	267	GAAACATCACTGAGAA-----AATGCCAAGCCTGAAGTTGCA	305
QY	144	LysThrglyLeuglySerSeraglyLeuValSerValAlaAlaArhrSerLeuSer	163
Db	306	AAGACTGGATTAGGGTCATCAGCAGCATGACCACTCAGCTTGGCCGCTTCTCAT	365
QY	164	HsPhe---lIeProAnValIleSerThrsAnlys-----	174
Db	366	TATCTTGCGCTGTGAACCTCTCTCTTTGGATGAGACCACTCAAGGAAGAAAGAT	425
QY	175	-----AspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAlaGln	190
Db	426	CTACTGATCTGTGATGTAGCTCATGTGATGCTCAAACTGCTATTCATTCGACAGGT	485
QY	191	LysIleeglySerGlyPheAspValAlaThrAlaIleTyrglyLeuIleValTyrArg	210
Db	486	AAAGTTGCGAGTGCGTTGATGTAGTTCGACGTTATGGAAGTCAACGCTATATCCG	545
QY	211	PheGlnProAlaLeuIleasn	217
Db	546	TTTTCGCTGAAGTCTCTCT	566
RESULT 38			
LOCUS	BE494575	513 bp	mRNA
DEFINITION	WH1255_P10_K19S Secale cereale anther cDNA library Secale cereale		
ACCESSION	BE494575		
VERSION	BE494575.1	GI:9661168	
KEYWORDS	EST.		
SOURCE	Secale cereale (rye)		
ORGANISM	Secale cereale		
REFERENCE	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae		
AUTHORS	1 (bases 1 to 513)		
TITLE	Anderson, O.D., Butler, E., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Gustafson, J.P., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rauschen, C.J., Rose, K., Setton, C.L. and Tong, J.C.		
JOURNAL	The structure and function of the expressed portion of the wheat genomes - Anther cDNA library from rye		
COMMENT	Unpublished		
FEATURES	Feature: Olin Anderson		
SOURCE	US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center		
	800 Buchanan Street, Albany, CA 94710, USA		
	Tel: 5105595773		
	Fax: 5105595818		
	Email: oanders@pw.usda.gov		
	Sequence has been trimmed to remove vector sequence and low quality sequence with phred score less than 20		
	Seq primer: StrataGene SK primer.		
	Location/Qualifiers		
	1..513		
	/organism="Secale cereale"		
	/mol_type="mRNA"		
	/culivar="Blanco"		
	/db_xref="taxon:4550"		
	/clone="WH1255_P10_K19"		
	/tissue_type="Anther"		
	/dev_stage="Adult plant before anthesis"		
	/lab_host="B. coli SOLR"		
	/clone_lib="Secale cereale anther cDNA library"		
	/note="Vector: Lambda Uni-ZAP XR, excised phagemid;		
	Site 1: EORI; Site 2: XhoI; Plants were grown in the greenhouse. Anthers were harvested and pooled from early meiosis to late meiosis. The tissue, total RNA, and poly(A) RNA were prepared (Butler, Rose and Gustafson) at University of Missouri, Columbia. A cDNA library was made, and the cDNA clones were in vivo excised to give		

ORGANISM

Triticum aestivum

REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticaceae; Triticum.

AUTHORS

Tingey, S.V., Powell, W., Wolter, P., Dolan, M., Hainey, C., Yuan, Z.,

TITLE

Miao, G., Caraher, N. and Hanafey, M.K.

JOURNAL

DuPont Wheat cDNA Sequence

COMMENT

Unpublished

CONTACT

Contact: Scott V. Tingey

Crop Genetics

E. I. DuPont de Nemours and Company

1 Innovation Way,

P.O. Box 6104, Newark, DE 19714-6104, USA

Tel: 302-631-2607

Fax: 302-631-2607

Email: Scott.V.Tingey@USA.dupont.com

Seq primer: M13.

Location/Qualifiers

1. 515

/organism="Triticum aestivum"

/mol_type="mRNA"

/cultivar="Stephens"

/db_xref="taxon:4555"

/clone="wlm96.pk027.e4"

/issue_type="leaf"

/clone_lib="wlm96"

/note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:

XhoI; Wheat (Triticum aestivum L.) seedlings 96 hr after

inoculation w/ E. graminis"

BASE COUNT

104 a 170 c 105 g 135 t 1 others

ALIGNMENT

Scores:

Pred. No.:

6.81e-09

Score:

173.00

Percent Similarity:

45.73%

Best Local Similarity:

29.88%

Query Match:

7.76%

DB:

14

Gaps:

4

US-10-069-062-7 (1-432) x CA684704 (1-515)

QY

19 ValLeuGluProIleTyAspAlaTyValThrAlaLeuSerSerArgMetHisAlaVal 38

DB

2 GTGCTTGACCGCTGACACCGCGCTTGTGCTGACGCGCGCATTCATTCATTC 61

QY

39 ILeThrProLyseGlyThrsSerLeuys-----GluSerArgIleLysIleSerSerPro 56

DB

62 GTGCTGACCGCGCTGACACCGCGCTTGTGCTGACGCGCGCATTCATTCATTC 121

QY

57 GluPheAlaAsnGlyGluTrpGluTyHisIleSerSerAsnThrGluysProArgIu 76

DB

122 CAGTTCCTGAGAGCTCAATGGGATATGACGCCAGTGCATGCGCAACAATCGCGCTT 181

QY

77 ValGlnSerArgIle-----AsnProPheLeuGlu 86

DB

182 ATCGCTCTCTTGTAGATCCAGCCCTACGCGCGAGCAGTCCGCAATCTGTATTGAA 241

QY

87 AlaThrIlePheIleValLeuAlaTyIle-----GlnProThrGluAlaPhe 102

DB

242 ACTACACTGACATATGATTGACTTACATCGCCGAACCTGTCCACAACTACATCATTC 301

QY

103 AspleuGluIleIleIleTySerAspProGlyTyHisSerGlnGluysPthrGluThr 122

DB

302 CATGCCAGCATCATTCCTTCCGATGCGATTAATCTCG-----AATCCA 349

QY

123 LysThrSerSerAsnGlyGluysThrPheLeuTyHisSerArgAlaIleThrGluVal 142

DB

350 CCGTATCCCTGACCTGCTATCTGCTTCAAACTTCTGTCCCGCTGCGAGTCTCT 409

QY

143 GluysThrGlyLeuGlySerSerAlaGlyLeuValSerValAlaIleThrSerLeu 162

DB

410 CACAAGACTGAGCTTGGCTCATGCTGCTGAGCGCGCTTGAATGCGCTCTCTCTC 469

QY 163 SerHisPheIle 166
DB 470 AACCACTATCTA 481Search completed: August 15, 2003, 16:28:52
Job time : 2829 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 15, 2003, 15:56:39 / Search time 320 Seconds
(without alignments)
3021.219 Million cell updates/sec

Title: US-10-069-062-7
Perfect score: 2230
Sequence: 1 MSKAFSAPKAFKLAGYLV.....DLEQTEGVLEKPEPDYIGL 432

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Delop 6.0, Delext 7.0

Searched: 1504479 seqs, 1118970152 residues
Total number of hits satisfying chosen parameters: 3008958

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n model
-g=/cgn2_1/USPTO_spool/US10069062/runat_07082003_161506_4284/app_query.fasta_1.583
-DB=PubMed_Applications_NA_QMFT-faastap -SUFFIX=rmpb -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pcp -THR MAX=500
-THR MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=pcp -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USRR=US10069062 @cgn_1.1.221 @runat_07082003_161506_4284
-NCPUB=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09D_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	809.5	36.3	1356 11	US-09-918-740-48 Sequence 48, Appl

2	809.5	36.3	1356	12	US-10-006-909-5	Sequence 5, Appl
3	809.5	36.3	4482	12	US-10-006-909-9	Sequence 9, Appl
4	809.5	36.3	5051	12	US-10-006-909-12	Sequence 12, Appl
5	809.5	36.3	5963	12	US-10-006-909-13	Sequence 13, Appl
6	809.5	36.3	7681	11	US-09-918-740-61	Sequence 61, Appl
7	809.5	36.3	7693	11	US-09-918-740-58	Sequence 58, Appl
8	809.5	36.3	7695	11	US-09-918-740-52	Sequence 52, Appl
9	809.5	36.3	8224	11	US-09-918-740-60	Sequence 60, Appl
10	809.5	36.3	8235	11	US-09-918-740-64	Sequence 64, Appl
11	809.5	36.3	8400	11	US-09-918-740-60	Sequence 64, Appl
12	809.5	36.3	9253	12	US-10-006-909-7	Sequence 7, Appl
13	809.5	36.3	13917	11	US-09-918-740-72	Sequence 72, Appl
14	809.5	36.3	14623	11	US-09-918-740-74	Sequence 74, Appl
15	809.5	36.3	14623	11	US-09-918-740-76	Sequence 76, Appl
16	671	30.1	1356	14	US-10-128-714-2406	Sequence 2406, Ap
17	655.5	29.4	1455	14	US-10-128-714-7406	Sequence 7406, Ap
18	648.5	29.1	1509	14	US-10-128-714-1406	Sequence 1406, Ap
19	649.5	29.1	1509	14	US-10-128-714-6406	Sequence 6406, Ap
20	649.5	29.1	3508	14	US-10-128-714-406	Sequence 406, Ap
21	649.5	29.1	3509	14	US-10-128-714-5406	Sequence 5406, Ap
22	475	21.3	2396	10	US-09-988-863A-1	Sequence 1, Appl
23	445.5	20.0	1509	14	US-10-036-859B-5	Sequence 5, Appl
24	231	10.4	728	10	US-09-988-863A-4	Sequence 4, Appl
25	213	9.6	5158	14	US-10-128-714-4038	Sequence 4038, Ap
26	209.5	9.4	1107	9	US-09-815-242-6559	Sequence 6559, Ap
27	209.5	9.4	6835	10	US-09-070-927A-204	Sequence 204, Ap
28	181.5	8.1	757	10	US-09-909-745-21	Sequence 21, Appl
29	173	7.8	1035	9	US-09-815-242-4799	Sequence 8799, Ap
30	173	7.8	1077	9	US-09-815-242-8768	Sequence 8768, Ap
31	173	7.8	10146	8	US-08-781-986A-243	Sequence 243, Ap
32	147	6.6	6798	11	US-09-918-740-57	Sequence 57, Appl
33	147	6.6	8077	11	US-09-918-740-63	Sequence 63, Appl
34	140.5	6.3	539	10	US-09-909-745-19	Sequence 19, Appl
35	124.5	5.6	571	10	US-09-988-863A-5	Sequence 5, Appl
36	124.5	5.6	1332	11	US-09-918-740-18	Sequence 18, Appl
37	124.5	5.6	1332	11	US-09-918-740-49	Sequence 49, Appl
38	124.5	5.6	1332	12	US-10-006-909-4	Sequence 4, Appl
39	123	5.5	1137	10	US-09-938-842A-1231	Sequence 1231, Ap
40	122	5.5	1008	9	US-09-815-242-9572	Sequence 9572, Ap
41	116.5	5.2	443	9	US-09-770-444-851	Sequence 851, Ap
42	115	5.2	611	10	US-09-988-863A-3	Sequence 3, Appl
43	111	5.0	3750	10	US-09-938-842A-1662	Sequence 1662, Ap
44	109	4.9	3727	13	US-10-026-188-4	Sequence 4, Appl
45	109	4.9	4157	9	US-09-834-792-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-918-740-48
Sequence 48, Application US/09918740
Publication No. US20030033626A1
GENERAL INFORMATION:
APPLICANT: Hahn, Frederick
TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways t
FILE REFERENCE: KAS-103XCI
CURRENT FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: 60/221,703
PRIOR FILING DATE: 2000-07-31
NUMBER OF SEQ ID NOS: 76
SOFTWARE: PatentIn version 3.0
SEQ ID NO 48
LENGTH: 1356
TYPE: DNA
ORGANISM: Saccharomyces cerevisiae
US-09-918-740-48
Alignment Scores: 1.13e-83 Length: 1356
Pred. No.: 809.50 Matches: 190
Score:

Percent Similarity: 57.70% Conservative: 76
 Best Local Similarity: 41.21% Mismatches: 146
 Query Match: 36.30% Indels: 49
 DB: 11 Gaps: 14

US-10-069-062-7 (1-432) x US-09-918-740-48 (1-1356)

```

QY 3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyLysLeuValLeuGluPro 22
DB 13 AGAGCTTCAGTCCCGGAGGAAAGCTTACCTGCTGCTGATATTTAGTTAGATACA 72
QY 23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys 42
DB 73 AAATAGAAGCATTTGTAGTCGATATCGCAAGAATGATGCTGTAGCCCATCTTAC 132
QY 43 GlyThrSerLeuLysGluSer-----ArgIleLysIleSerSerProGlnPheAla 59
DB 133 GGT--TCATTTGCAAGGGTCTGATTAAGTTGAAGTGGTGTGAAAGTAAACAATTAA 189
QY 60 AsnGlyLysTrpGluTyrHisIleSerSerSerSerThrGlu--LysProArgGluValGln 78
DB 190 GATGGGAGAGTGGCTGATACCATTAAGTCTTAAAGTGGCTTCTCTGTTGATAGGC 249
QY 79 SerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnPro 98
DB 250 GATCTAAGAACCTTTTCATTTGAAAAAAGTTATCGCTAAGCTATTAGCTATTAAACCT 309
QY 99 Thr-----GluAlaPheAspLeuGluIleIleIleTyrSerAsp 111
DB 310 AACATGACGACTACTGCAATAGAACTGTTGTTATTTAT--ATTTCTCTGAT 363
QY 112 ProGlyTyrHisSerGlnGluAspThrGluThrLysThrSerSerAsnGlyLysTyr 131
DB 364 GATGCTTACCATTTCTGAGAGGATACGCTTACCGAA-----CATGCTGGCAACAGAA 417
QY 132 PheLeuTyrHisSerArgAlaIleThrGluValGlnLysThrGlyLeuGlySerSerAla 151
DB 418 TTGAGTTTTCATTCGACAGAAATTTGAAGAAATTTCCAAAAACAGGCTGGCTCGGCA 477
QY 152 GlyLeuValSerValAlaThrSerLeuSerIlePheIle-----ProAla 168
DB 478 GGTATGTCACAGTTTAACTACAGCTTGGCTCTTTTGTATTCGACCTGGAATAAT 537
QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
DB 538 AATGTAGACAAATATAGAGAAATTTATTCATATTAGCACAATGCTGCTATTCAGCT 597
QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
DB 598 CAGGCTAAAAATTTGGAAGCGGTTGATGTAGCGCGGACAGCATATGATCATCAGATAT 657
QY 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGlu 228
DB 658 AGAAGATTCACACCCCATTAATCTTAATTTGCCAGATAT-----GGAGTGGCT 708
QY 229 LysPheProThrGluLeuLysLysLeuIle--GluSerAsnTrpGluGluLysHisGlu 247
DB 709 ACTTACGCGAGTAACTGGCGGCTTGGTTGATGAAGAAAGACATGCAATATTACGATTAA 768
QY 248 ArgCysThrLeuProTyrGlyLysLeuLeuMetGlyAspValLysGlyLysGlu 267
DB 769 AGTAACCATTTACCTTCGGGATTAATCTTATGATGCGGCAATTTAAGATGTTACAGA 828
QY 268 ThrProLysLeuValSerArgValLeuGlnTrpLysGlyLysProGluGluSerSer 287
DB 829 ACAGTAAACTGGTCCAGAGGTAATAAATTGGTATGATTCGATATGCCAGAACTCTG 888
QY 288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
DB 889 AAAATATATACAACTCATCATGCAATTTCTAGATTTATGATGATGATCTTAAACAT 948
QY 305 ArgGluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHis 324
DB 305 ArgGluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHis 324

```

```

DB 949 GATCGCTTACACGAGACTCATGACGATTACAGCGATCATGATATTTAGTCTTACAGAG 1008
QY 325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
DB 1009 AATGACTGTACTCTGCAAAAAGTATCTGAAATATCAGAAAGTTAGAGATGCACTGCCACA 1068
QY 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspVal 355
DB 1069 ATTAGACGTTCTCTTAGAAAAAATACTAAGAAATCTGGTGGCCGATATGAACTCCCGTA 1128
QY 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
DB 1129 CAACCTAGCTTATTTGATGATGATGATCCAGACTTAAAGAGATTTCTTACTTAAATACCT 1188
QY 376 GlyAlaGlyLysTyrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
DB 1189 GGTGCTGGGTGATTAAGACCCATTCAGATGATTAACGAAAGATGTTATGAGGCT 1248
QY 391 GlnValGlyAsnProPheLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
DB 1249 CAACCGCTAATGACAAAAGA-----TTTTCTAAGGTTCA 1284
QY 411 TrpValAspLeuGluGlnGlnThrGluGlyValLeuGluGluLys--ProGluAspTyr 429
DB 1285 TGGCTGATGTAACTCAGGCTGATCGGGGTGTTAGAAAAAAGATCCGGAACCTTAT 1344
QY 430 Ile 430
DB 1345 CTT 1347

```

RESULT 2
 US-10-069-909-5
 ; Sequence 5, Application US/10006909
 ; Publication No. US20030148479A1
 ; GENERAL INFORMATION:
 ; APPLICANT: KEASLING, JAY
 ; APPLICANT: MARTIN, VINCENT
 ; APPLICANT: PITERA, DOUGLAS
 ; APPLICANT: KIM, SEON-WON
 ; APPLICANT: WITHERS III, SYDNOR T.
 ; APPLICANT: YOSHIKUNI, YASUO
 ; APPLICANT: NEWMAN, JACK
 ; APPLICANT: KHEBNIKOV, ARTEM VALENTINOVICH
 ; TITLE OF INVENTION: BIOSYNTHESIS OF ISOPENTENYL PYROPHOSPHATE
 ; FILE REFERENCE: 2000-0007
 ; CURRENT APPLICATION NUMBER: US/10/006,909
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 5
 ; LENGTH: 1356
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; OTHER INFORMATION: Phosphomethylalate kinase nucleotide sequence
 US-10-069-909-5

Alignment Scores:
 Pred. No.: 1,136-83 Length: 1356
 Score: 809.50 Matches: 190
 Percent Similarity: 57.70% Conservative: 76
 Best Local Similarity: 41.21% Mismatches: 146
 Query Match: 36.30% Indels: 49
 DB: 12 Gaps: 14

US-10-069-062-7 (1-432) x US-10-069-909-5 (1-1356)

```

QY 3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyLysLeuValLeuGluPro 22
DB 13 AGAGCTTCAGTCCCGGAGGAAAGCTTACCTGCTGCTGATATTTAGTTAGATACA 72
QY 23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys 42
DB 73 AAATAGAAGCATTTGTAGTCGATATCGCAAGAATGATGCTGTAGCCCATCTTAC 132

```

```

Db      73  AATATGAAGCACTTTAGTCGATTAATCGCAAGATGATGCTAGCCCATCTTAC 132
      43  G1YThrSerLeuYsgUser-----Arg1leYs1leSerSerProGlnPheA 59
      133  GGT---TCATTGCAAGGCTGTGATTAAGTTGAAGCTGTGTAAGAAAGTAAACATTAA 189
      60  AsnG1gluTrpGluTrpHis1leSerSerSerThrGlu---LysProArgGluValGln 78
      190  GATGGGAGTGGCTGATACCATTAATCTCTAAAGGCGCTCATCTCTGTCGATAGGC 249
      79  SerArg1leAsnProPheLeuGluValatThr1lePhe1leValleuValTyr1leGlnPro 98
      250  GGATCTAAGAACCTTTCACTGTAAGAAAGTATGCTAAGCTATTGACTTAAACCT 309
      99  Thr-----Glu1lePheAspLeuGln1le1leTyrSerAsp 111
      310  AACATGAGACGACTCTGCAATGAAACCTGTTGCTTATGAT---ATTTCCTGTAT 363
      112  ProG1YThrHisSerGlnGluAspThrGluTrpHisThrSerSerAsnG1gluYstr 131
      364  GATGCTTACCATCTTCAGAGGATAGCTTACCGA-----CATGCTGGCAACAGAA 417
      132  PheLeuTrpHisSerArg1leThrGluValGluYstrG1YleuG1YserSerAla 151
      418  TTGAGTTTCATTCGACAGAAATGGAAGATTCCCAAAACAGGCGCTGCTCGGCA 477
      152  G1YleuValSerValValAlaThrSerLeuLeuSerHisPhe1le-----ProAsn 168
      478  GGTATGATGACAGTTTAACTACAGCTTTGGCTCTTTTGTATGCGACCTGAAAT 537
      169  Val1leSerThrAsnLysAsp1leLeuHisAsnValAlaGln1leAlaHisCysTyrAla 188
      538  AATGTAGCAAAATATAGAGAAGTATCTAATATTAGCAACAGTCTCTATGTCAGCT 597
      189  GlnLysLys1leG1YserG1YpheaSerValAlaThrAla1leTyrG1Yleu1leValTyr 208
      598  CAGGGTAAATGGAAGCGGCTTGTATGATGCGCGGCGACATGATGATCTACAGATAT 657
      209  ArgArgPheGlnProAlaLeu1leAsnAspValPheGlnValleuGlnSerAspProGlu 228
      658  AAGAAGTTCCACCGCATTAATCTCTAATTTGCGAGATAT---GGAAGTCT 708
      229  LysPheProThrGluLeuLysLysLeu1le---GluSerAsnTrpGluGlnLysHisGlu 247
      709  ACTTACGCGAGTAACTGCGCATTTGCTGATGAAGAAAGACTGCAATTTACGATTAA 768
      248  ArgCysThrLeuProTyrG1YleLysLeuLeuMetG1YAspValLysG1YserGlu 267
      769  ACTTACCATTTACCTTCGGGATTAACCTTATGATGCGGATTAAGAAAGTGTCAAGA 828
      268  ThrProLysLeuValSerArgValleuGlnTrpLysLysGluYstrProGlnLysSer 287
      829  ACGATTAACCTGTCGCAAGAGTAAATAATGCTATGATTCGCATATGCCAAGGCTTG 888
      288  ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGlnLeu 304
      889  AAAATATATACAGACTGCATGCAATTTCTAATTTATGATGAGCATATCTTAAC 948
      305  ArgGluMetArgGluLysTyrAspSerAspProGluTrpTyr1leYsg1leLeuAspHis 324
      949  GATCCCTTACAGACATCATGATGATACACGCGATCATATTTAGAGCTCTTGAGAG 1008
      325  Ser-----ValGluProLeuThrValAla1leLysAsn 335
      1009  AATGACTGACTGTCAAAAAGTATCTGTAATCAAGAACTTGAAGATGCAATGCCACA 1068
      336  IleArgLysG1YleuGlnAlaLeuThrGlnLysSerGluValPro1leGluProAspVal 355
      1069  ATTAGACGTTCTTATAGAAAATTAATAAGATCTGCGCGGATATGCAACTCCCGTA 1128
      356  GlnThrGlnLeuLeuAspArgCysGlnGln1leProG1YCysValG1YglYValValPro 375

```

```

Db      1129  CAAACTAGCTTATTTGATGATGTTGCCAGACTTAAAGAGGTTCTTACTGTTAACT 1188
      376  G1YAlaG1YglYTrpAspAla1leAlaValleuVal-----LeuGlnAsn 390
      1189  GGTGCTGGTGTATATGAGCCCATTCAGATGATTACTTAAGCAAGATGTTGATTTAGGCT 1248
      391  GlnValG1YAsnPheLysGlnLysThrLeuGlnAsnProAspTyrPheHisAsnValTyr 410
      1249  CAACCGCTAATGACAAAAGA-----TTTTCTAAGCTTCA 1284
      411  TrpValAspLeuGluGlnGlnThrGluG1YValleuGlnGluYstr---ProGluAspTyr 429
      1285  TGGCTGATGTAACTCAGAGCTGATCGGGGTGTATAGAAAGAAAGATCCGAACTTAT 1344
      430  Ile 430
      1345  CTT 1347

RESULT 3
US-10-006-909-9
; Sequence 9, Application US/10006909
; Publication No. US20030148479A1
; GENERAL INFORMATION:
; APPLICANT: KEASLING, JAY
; APPLICANT: MARTIN, VINCENT
; APPLICANT: PITERA, DOUGLAS
; APPLICANT: KIM, SEON-WON
; APPLICANT: WITHERS III, SYDOR T.
; APPLICANT: YOSHITOMI, YASUO
; APPLICANT: NEWMAN, JACK
; APPLICANT: KHEBNIKOV, ARTEM VALENTINOVICH
; TITLE OF INVENTION: BIOSYNTHESIS OF ISOPENTENYL PYROPHOSPHATE
; FILE REFERENCE: 2000-0007
; CURRENT APPLICATION NUMBER: US/10/006,909
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 9
; LENGTH: 4482
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-006-909-9

Alignment Scores:
Pred. No.: 7.35e-83 Length: 4482
Score: 809.50 Matches: 190
Percent Similarity: 57.70% Conservative: 76
Best Local Similarity: 41.21% Mismatches: 146
Query Match: 36.30% Indels: 49
DB: 12 Gaps: 14

US-10-069-062-7 (1-432) x US-10-006-909-9 (1-4482)
      3  LysAlaPheSerAlaProG1YValAlaPheLeu1leG1YTrpLeuValleuGluPro 22
      1610  AGAGCTTCAAGTGGCCCGGAGAAAGCTTACTGCTGTGATTAATTTAGATTGATCA 1669
      23  IleTyrAspAlaTyrVal1leAlaLeuSerSerArgMetHisAlaVal1leThrProLys 42
      1670  AATATGAGCACTTTGTATGTCGATTAATCGCAAGATGCAATGCTGTAGCCCATCTTAC 1729
      43  G1YThrSerLeuYsgUser-----Arg1leYs1leSerSerProGlnPheA 59
      1730  GGT---TCATTGCAAGGCTGTGATTAAGTTGAATGCGGTGGAAGAAAGTAAACATTAA 1786
      60  AsnG1gluTrpGluTrpHis1leSerSerSerThrGlu---LysProArgGluValGln 78
      1787  GATGGGAGTGGCTGATACCATTAATCTCTAAAGGCGCTCATCTCTGTCGATAGGC 1846
      79  SerArg1leAsnProPheLeuGluValatThr1lePhe1leValleuValTyr1leGlnPro 98

```



```

Db      2195 CAGGTAATAATGGAAGCGGGTTGATGTAGCGCGGACGATGATGATCTATCAGATAT 2254
QY      209 ArgArgPheGlnProAlaLeuIleAenAspValPheGlnValLeuGluSerAspProGlu 228
Db      2255 AGAAGATTCACCCCGCATTAATCTTAATTTCCAGATAT-----GGAAGTCT 2305
QY      229 LysPheProThrGluLeuLysLeuLeu-----GluSerAsnTrpGluGluShiGlu 247
Db      2306 ACTTAGCGGAGTAAGTAAGTGGCGCATTTGGTATGAGAAAGACTGGAAATATAGATTAA 2365
QY      248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetCysAspValLysGlySerGlu 267
Db      2366 AGTAACCATTTACTTCGGGATTAATCTTAATGATGAGCGCATATTAAGATGGTTACAGA 2425
QY      268 ThrProLysLeuValSerArgValLeuGlnTrpLysGluLysProGluGluSer 287
Db      2426 ACGTAATAATGGTGTCCAGAGGTAAATAATGGTATGATTCGATATGCCAGAAACCTTG 2485
QY      288 ValValTyrAspGlnLeuAenSerAlaAenLeuGlnPheMet-----LysGluLeu 304
Db      2486 AATAATATATACAACTCCATCATGCAAAATTTAGATTATGATGATGATGATGATGATGAT 2545
QY      305 ArgGluMetArgGlyLysTyrAspSerAspProGluTrpTyrIleLysGluLeuAspHis 324
Db      2546 GATCGCTTACACAGACTCATGACGATTCAGCGATCATGATATTGATGATCTTTGAGAGG 2605
QY      325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
Db      2606 AATGACTGATACCTGCTCAAAAGTATCCTGAATATCACAGAGTGAAGATGAGTGCACAC 2665
QY      336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspVal 355
Db      2666 ATTAGACGTTCTTTGAAATAATATCAAAAGATCTGTCGCCGATATCGAACCTCCCGTA 2725
QY      356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
Db      2726 CAACACTGACTTATGTGATGATTCACAGACCTTAAAGAGTCTTACTTCTTAATAACCT 2785
QY      376 GlnValGlyLysTyrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
Db      2786 GGTGCTGGTGGTATATGACGCGCATTCGATGATTAAGCAAGATGTTGATTCAGCGCT 2845
QY      391 GlnValGlyAsnAspLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
Db      2846 CAACCGCTAATACAAAAGA-----TTTCTAGGTTCA 2881
QY      411 TrpValAspLeuGluGlnGlnThrGluGlyValLeuGluGluLys---ProGluAspTyr 429
Db      2882 TGGCTGATGTAATCACTGAGCTGATGGGTGTGTAGGAAAGAAAGATCCGAAACTTAT 2941
QY      430 Ile 430
Db      2942 CTT 2944

```

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Operon C containing A. thaliana, S. cerevisiae, and Streptomyces
; OTHER INFORMATION: C1.90
; OTHER INFORMATION: DNA, and R. capsulatus DNA
US-09-918-740-61

Alignment Scores:
Score: 1,716-82 Length: 7681
Percent Similarity: 809.50 Matches: 190
Best Local Similarity: 57.70% Conservative: 76
Query Match: 41.21% Mismatches: 146
DB: 36.30% Indels: 49
Gaps: 14

US-10-069-062-7 (1-432) x US-09-918-740-61 (1-7681)
QY      3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyLysTyrLeuValLeuGluPro 22
Db      47 AGAGCCTTCAGTCCCGGAGGAAAGCGTTACTAGCTGGTGGATTTAGTTAGATACA 106
QY      23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys 42
Db      107 AATATGAAGCATTTGTATGTCGATTAATCGGCAAGAAAGCATGTCGTAGCCCATCTTAC 166
QY      43 GlyThrSerLeuLysGluSer-----ArgIleLysIleSerSerProGlnPheAla 59
Db      167 GGT---TCATTTCAGAGGCTCTGATATGATTTGAAGTGGCGGTGCAAAAGTAAACATTTAA 223
QY      60 AsnGlyLysTrpGluLysThrIleSerSerAsnThrGlu---LysProArgGluValGln 78
Db      224 GATGGGAGATGCTGATGATTAAGTCTTAAGTGGCTTCATTCCTGTTTCGATAGGC 283
QY      79 SerArgIleAsnProPheLeuGlnAlaThrIlePheIleValLeuAlaTyrIleGlnPro 98
Db      284 GGATCTAAGAACCTTTTCATTTGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 343
QY      99 Thr-----GluAlaPheAspLeuGluIleIleTyrSerAsp 111
Db      344 AACATGAGACGACTACTGCAATGAAACCTTTCCTTATTTAGAT-----ATTTCCTCAT 397
QY      112 ProGlyTyrHisSerGlnGluAspThrGluThrLysThrSerSerAsnGlyLysLysThr 131
Db      398 GATGCTTACATTCACAGAGATGATGCTTACCGAA-----CATCGTGCACAGAGA 451
QY      132 PheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAla 151
Db      452 TTGAGTTTCATTCGACAGATTAAGAAAGTTCACAAACAGGCGCTGGCTCCCGCA 511
QY      152 GlyLeuValSerValValAlaThrSerLeuLeuSerHisPheIle-----ProAsn 168
Db      512 GGTTAGTCACAGTTTATCTACAGCTTTGGCTTCCTTTTGTATCGGACCTGGAAT 571
QY      169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
Db      572 AATGTAGCAATATATAGAGAGATTAATCATATTTAGCAAGATGTCATTTGCAAGCT 631
QY      189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
Db      632 CAGGTAATAATGGAAGCGGGTTGATGTAGCGCGGACGATATGATGATGATGATGATGAT 691
QY      209 ArgArgPheGlnProAlaLeuIleAenAspValPheGlnValLeuGluSerAspProGlu 228
Db      692 AGAAGTTCACCCCGCATTAATCTTAATTTCCAGATAT-----GGAAGTCT 742
QY      229 LysPheProThrGluLeuLysLeuLeu-----GluSerAsnTrpGluGluShiGlu 247
Db      743 ACTTAGCGGAGTAAGTAAGTGGCGCATTTGGTATGAGAAAGACTGGAAATATAGATTAA 802
QY      248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetCysAspValLysGlySerGlu 267
Db      803 AGTAACCATTTACTTCGGGATTAATCTTAATGATGAGCGCATATTAAGATGGTTACAGA 862

```



```

QY 268 ThrProlyseuValSerArgValLeuGlnTrpLysGluLysProGluSerSer 287
Db 863 ACAGTAAACTGTCAGAGGTAATAAATTGTATGATTCGCATATGCAGAAAGCTTG 922
QY 288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
Db 923 AAAATATATACAGAACTGATCATGCATAATTTAATTTATGATGACATCTATCTAACTA 982
QY 305 ArgGluMetArgGluLysTyrAspSerAspProGluTrpTyrLysGluLeuAspHis 324
Db 983 GATCGCTTACACAGACATCATACATTAACAGCGATCAGATTTTGAAGCTCTTGAAGAG 1042
QY 325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
Db 1043 AATGACTGACCTGTCAAAAGATATCTGAATACACAGAAATGAGATGACAGTCCACA 1102
QY 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspVal 355
Db 1103 ATTAGACGCTTCTTTAGAAAATACTAAGATCTGTCGCCGATATGAACTCCCGTA 1162
QY 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
Db 1163 CAACTAGCTTATTTGATGATGATTCAGACCTTAAAGAGTTCTTACTTCTTAACT 1222
QY 376 GlyAlaGlyGlyTyrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
Db 1223 GGTGCTGGGTGTTATGACGCCATTCAGATGATTAACAAGATGATGATCTTGGCT 1282
QY 391 GlnValGlyAsnPheLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
Db 1283 CAACCGCTAATGACAAAGA-----TTTTCTAAGCTTCA 1318
QY 411 TrpValAspLeuGluGluGlnThrGluGlyValLeuGluLys---ProGluAspTyr 429
Db 1319 TGCGTGATGTTACTCAGCGCTGACTGGGTGTTAGGAAGAAAGATCCGAACTTAT 1378
QY 430 Ile 430
Db 1379 CTT 1381

```

```

RESULT 7
US-09-918-740-58
; Sequence 58, Application US/09918740
; Publication No. US20030033626A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
; FILE REFERENCE: KAS-103XCI
; CURRENT APPLICATION NUMBER: US/09/918,740
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221,703
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 58
; LENGTH: 7693
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Operon containing A. thaliana and S. cerevisiae DNA
US-09-918-740-58

```

```

Alignment Scores:
Pred. No.: 1.71e-82 Length: 7693
Score: 809.50 Matches: 190
Percent Similarity: 57.70% Conservative: 76
Best Local Similarity: 41.21% Mismatches: 146
Query Match: 36.30% Indels: 49
DB: 11 Gaps: 14

```

```

US-10-069-062-7 (1-432) x US-09-918-740-58 (1-7693)
QY 3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyTyrLeuValLeuGluPro 22
Db 3800 AGAGCTTCACTGCCCCAGGAAAGCGTTACTAGCTGGTGAATTTTATGTTTAAATACA 3859
QY 23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys 42
Db 3860 AAAATATGACGATTTGTGTCGATTAATCCGCAAGAAATGATGCTGTAGCCATCTTAC 3919
QY 43 GlyThrSerLeuLysGluSer-----ArgIleLysIleSerSerProGluPheAla 59
Db 3920 GGT---TCATTGCAAGGCTGTGATTAAGTTGAAGCTGTGAAAAGTAAACAATTTTAA 3976
QY 60 AsnGlyLysTrpGluTyrThrIleIleSerSerAsnThrGlu---LysProArgLysValGln 78
Db 3977 GATGGGAGTGGCTGTACATTAATGCTTAAAGTGCTTCATTCCTGTTCCATAGGC 4036
QY 79 SerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnPro 98
Db 4037 GGATCTAAGAACCTTTCATTGAAAAGTTATCCCTAAGTATTAGTACTTAAACCT 4096
QY 99 Thr-----GluAlaPheAspLeuGluIleIleLysSerAsp 111
Db 4097 AACATGACGACTACGCAATAGAACTGTGCTTATTTGAT-----ATTTCTCTGAT 4150
QY 112 ProGlyTyrHisSerGlnGluAspThrGluThrLysThrSerSerAsnGlyLysThr 131
Db 4151 GATGCTTACATTTCTCAGAGAGATAGCTTACCGAA-----CATCGGCAACAGAA 4204
QY 132 PheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAla 151
Db 4205 TTGAGTTTCACTGCAACAGAAATGAAAGAGTCCCAAAACAGGCGTGGCTCCGCA 4264
QY 152 GlyLeuValSerValAlaIleThrSerLeuLeuSerHisPheIle-----ProAsn 168
Db 4265 GGTTTAGTACAGATTTTAATCTACAGCTTGGCTCTTTTGTATCGACCTGGAAT 4324
QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
Db 4325 AATGATAGCAAAATATAGAAAGTTATTCATTAATTTAGCAGAAAGTTCATTTGCAAGCT 4384
QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
Db 4385 CAGGTAATATGGAACGGGTTTATGATGTCGCGGAGCATATGATCATATGATAT 4444
QY 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGlu 228
Db 4445 AGAAGATTTCCACCGCATTAATCTTAATTTGCCAGATAT-----GGAAGTCT 4495
QY 229 LysPheProThrGluLeuLysLysLeuIle---GluSerAsnTrpGluLysHisGlu 247
Db 4496 ACTTACGCGAGTAACTGGCGCATTTGTTGATTAAGAAGCTGAATATTAACATTA 4555
QY 248 ArgCysThrLeuProTyrGlyLysLeuLeuMetGlyAspValLysGlyGlySerGlu 267
Db 4556 AGTACCATTTTACTCTGGGATTAATCTTAATGATGGCCATTAAGATGCTTCA 4615
QY 268 ThrProlyseuValSerArgValLeuGlnTrpLysGlyLysProGluGluSerSer 287
Db 4616 ACAGTAAACTGTCAGAGGTAATAAATTGTATGATTCGCATATGCCAGAAAGCTTG 4675
QY 288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
Db 4676 AAAATATATACAGAACTGATCATGCATAATTTAATTTATGATGACATCTATCTAACTA 4735
QY 305 ArgGluMetArgGluLysTyrAspSerAspProGluTrpTyrLysGluLeuAspHis 324
Db 4736 GATCGCTTACACAGACATCATTAACAGCGATCAGATTTTGAAGCTCTTGAAGAG 4795
QY 325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
Db 4796 AATGACTGACCTGTCAAAAGATCTGAATATCACAGAAATGAGATGAGTGGCCACA 4855

```

```

QY 336 ILeaRgLySgLYLeuGlAlaLeuThrGlnLySseRgLYValProIleGluProAspVal 355
DB 4856 ATTAGACGCTTCCTTAGAAAAAATACTAAAGATCTGCGCCGATTCACACTCCCGTA 4915
QY 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
DB 4916 CAAACTAGCTTAATGATGATGATGACACCTTAAAGAGATTCTTACTGCTTAATACCT 4975
QY 376 GYAlaGlyGlyTYrTrsPalaIleAlaValLeuVal-----LeuGluAsn 390
DB 4976 GGTGCTGGATGATATACCCGATTCGATGATTAACGACAGATGTTGATCTTACGGCT 5035
QY 391 GlnValGlyAsnPhelySgLYSthrLeuGluAsnProAspTYrPheHisAsnValTYr 410
DB 5036 CAACCCGCTAATACCAAAAGA-----TTTCTAAGCTTCA 5071
QY 411 TrpValAspLeuGlnGluGlnThrGlnGlyValLeuGlnGluLYs---ProGluAspTYr 429
DB 5072 TGGCTGGATGATTAACCTCAGGCTGACTGGGCTGTAGGAAAAAAGATCCGAAACTTAT 5131
QY 430 ILe 430
DB 5132 CTT 5134

RESULT 8
US-09-918-740-59
; Sequence 59, Application US/09918740
; Publication No. US20030033626A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; APPLICANT: Kuehnle, Adelheid
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
; FILE REFERENCE: KAS-103XCI
; CURRENT APPLICATION NUMBER: US/09/918,740
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221,703
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 59
; LENGTH: 7695
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Operon B containing A. thaliana and S. cerevisiae DNA
US-09-918-740-59

Alignment Scores:
Pred. No.: 1,72e-82 Length: 7695
Score: 809.50 Matches: 190
Percent Similarity: 57.70% Conservative: 76
Best Local Similarity: 41.21% Mismatches: 146
Query Match: 36.30% Indels: 49
DB: 11 Gaps: 14

US-10-069-062-7 (1-432) x US-09-918-740-59 (1-7695)
QY 3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyTYrIleuValLeuGluPro 22
DB 32 AGAGCCTTCAGTCCCGCAGGAAAGCGTTACTAGCTGCGATTAATTAGTTAGATACA 91
QY 23 IleTYrAspAlaTYrAlaThrAlaLeuSerSerArgMetHisAlaValIleThrProLYs 42
DB 92 AAATATGAAGCAATTGTAGTCGATTATCGGCAAGATGATGCTGTACCCATCTTAC 151
QY 43 GlyThrSerIleuLYSgLYser-----ArgIleLYsIleSerSerProGlnPheAla 59
DB 152 GGT---TCATTGCAAGGCTGATAGTTTGAAGTGCCTGTAAGAAACATTTAA 208
QY 60 AaNGlyGlyTrpGlyIuTYrHisIleSerSerAsnThrGlu---LysProArgGluValGln 78
DB 60 AaNGlyGlyTrpGlyIuTYrHisIleSerSerAsnThrGlu---LysProArgGluValGln 78

```

```

DB 209 GATGGGAGTGGCTGTACCATATATAGTCTTAAAGTGCCTTCATCTTGTTCATAGGC 268
QY 79 SerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTYrIleGlnPro 98
DB 269 GATCTTAAGAACCTTTCATTTGAAAAAAGTATGCGTAACGTAATTTAGCTTAAACT 328
QY 99 Thr-----GluAlaPheAspLeuGluIleIleTYrSerAsp 111
DB 329 AACATGACGACTACTGCATATGAAACTTGTTCGTTATGAT-----ATTTCCTCAT 382
QY 112 ProGlyTYrHisSerGlnGluAspThrGlnThrLYsThrSerSerAsnGlyGluLYsThr 131
DB 383 GATGCTTACCATTTCTCAGAGAGATGCGTTCGCA-----CATCGTGCAACAAAGA 436
QY 132 PheLeuTYrHisSerArgAlaIleThrGluValGluLYsThrGlyLeuGlySerSerAla 151
DB 437 TTGAGTTTCATTCACAGATTAAGAGATTTCCAAACAGGGCTGGCTCTCGGCA 496
QY 152 GlyLeuValSerValAlaIleThrSerLeuLeuSerHisPheIle-----ProAsn 168
DB 497 GGTTAGTCACAGTTTAACTACAGCTTTGGGCTCTTTTGTATCGGACCTGGAAT 556
QY 169 ValIleSerThrAsnLYsAspIleLeuHisAsnValAlaGlnIleAlaHisCysTYrAla 188
DB 557 AATGTAGACAAATATATAGAGAAATTATCATTAATTAGACAAAGTTCATTTGCATGCT 616
QY 189 GlnLYsLYsIleGlySerGlyPheAspValAlaThrAlaIleTYrGlyLeuIleValTYr 208
DB 617 CAGGCTAAATTTGGAAGCGGCTTATATAGGCGGCGACATATAGATCTTAATCAGATAT 676
QY 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGlu 228
DB 677 AGAATATTCACACCGCATTAATCTTAATTGCCAGATAT-----GGAAGTGT 727
QY 229 LysPheProThrGlnLeuLYsLYsLeuIle---GluSerAsnTrpGlnLYsIleGlu 247
DB 728 ACTTACGCGACAAACTGCGCATTTGGTTGATGAAGAAGCTGAATATTAACGATTAA 787
QY 248 ArgCysThrLeuProTYrGlyIleLeuSerLeuMetGlyAspValLYsGlySerGlu 267
DB 788 AGTAACCATTTACCTTCGCGATTAATCTTAATGATGCGGATATTAAGAAAGTTTCAGAA 847
QY 268 ThrProLYsLeuValSerArgValLeuGlnTrpLYsGluLYsProGlnGluSer 287
DB 848 ACAGTAAACTGCTCCAGAGCTAAATAATGATGATATTCGATATCCAGAAAGCTTG 907
QY 288 ValValTYrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LYsGluLeu 304
DB 908 AAATATATATACAGAACTGATCATGCAAAATCTTAATTTATGATGATGATCTTAACCTA 967
QY 305 ArgGluMetArgGluLYsTYrAspSerAspProGluThrTYrIleLYsGluLeuAspHis 324
DB 968 GATCCCTTAACAGACATCATGATGATTAACAGCGATCATTAATTTGATCTCTTAGAGG 1027
QY 325 Ser-----ValGluProLeuThrValAlaIleLYsAsn 335
DB 1028 AATGACTGATCCTGCAAAAGATATCCTGAATCAAGAAAGTTAGAGATGACGATGCCACA 1087
QY 336 ILeaRgLySgLYLeuGlAlaLeuThrGlnLYsSerGlyValProIleGluProAspVal 355
DB 1088 ATTAGACGCTTCCTTAGAAAAAATACTAAAGATCTGCGCGATGTAACCTCCGTA 1147
QY 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
DB 1148 CAAACTAGCTTAATGATGATGATGACCAAGCTTAAAGAGATTTACTTCTGCTTAATACCT 1207
QY 376 GYAlaGlyGlyTYrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
DB 1208 GGTGCTGGATGATAGAGCCATTCGATGATTAACGACAGATGTTGATGATCTTACGGCT 1267
QY 391 GlnValGlyAsnPhelySgLYSthrLeuGluAsnProAspTYrPheHisAsnValTYr 410
DB 1268 CAACCCGCTAATACCAAAAGA-----TTTCTAAGCTTCA 1303

```

QY 411 TTPValAspLeuGluGlnThrGluGlyValLeuGluGlyLys---ProGluAspTyr 429
 Db 1304 TGGCTGAGATGTAACACAGCTGACGCGGTGTAGGAAAGAAAGATCCGAAACTTAT 1363
 QY 430 Ile 430
 Db 1364 CTT 1366
 RESULT 9
 US-09-918-740-62
 ; Sequence 62, Application US/0918740
 ; Publication No. US20030033626A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hahn, Frederick
 ; APPLICANT: Kuehnle, Adelheid
 ; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
 ; TITLE OF INVENTION: create novel traits in transgenic organisms
 ; FILE REFERENCE: KAS-103XC1
 ; CURRENT APPLICATION NUMBER: US/09/918,740
 ; CURRENT FILING DATE: 2001-07-31
 ; PRIOR APPLICATION NUMBER: 60/221,703
 ; PRIOR FILING DATE: 2000-07-31
 ; NUMBER OF SEQ ID NOS: 76
 ; SOFTWARE: Patent version 3.0
 ; SEQ ID NO 62
 ; LENGTH: 8224
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Operon E containing A. thaliana, S. cerevisiae, Streptomyces sp. CI
 ; OTHER INFORMATION: DNA,
 ; OTHER INFORMATION: and R. capsulatus
 US-09-918-740-62
 Alignment Scores:
 Pred. No.: 1.9e-82 Length: 8224
 Score: 809.50 Matches: 190
 Percent Similarity: 57.70% Conservative: 76
 Best Local Similarity: 41.21% Mismatches: 146
 Query Match: 36.30% Indels: 49
 Db: 11 Gaps: 14
 US-10-069-062-7 (1-432) x US-09-918-740-62 (1-8224)
 QY 3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyGlyTyrLeuValLeuGluPro 22
 Db 47 AGAGCTTCAGAGCCCGGAGGAAAGCGTTACTAGCTGGTGGATTTTGGTTTATGATCA 106
 QY 23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys 42
 Db 107 AATATGAGAGCATTTGTAGTCGATTTATCGCAAAATGATGCTGTAGCCCATCTTAC 166
 QY 43 GlyThrSerLeuLysGluSer-----ArgIleLysIleSerSerProGluPheAla 59
 Db 167 GGT---TCATTGCAAGGCTTCGATTAAGTTTGAAGCGCTGCAAAAGTAAACAATTAA 223
 QY 60 AsnGlyLysTyrGluTyrHisIleSerSerSerThrGlu---LysProArgLysValGln 78
 Db 224 GATGGGAGGTGGCTGTACCAATTAAGTCTTAAAGTGGCTTCATTCCTGTTCCATAGGC 283
 QY 79 SerArgIleAsnProPheLeuGluAlaThrIlePheIleValIleAlaTyrIleGlnPro 98
 Db 284 GGATCTAGAAACCTTTTCATTCGAAAAGATATCGTAAACGTTTATAGCTATTAACCT 343
 QY 99 Thr-----GluAlaPheAspLeuGluIleIleTyrSerAsp 111
 Db 344 AACATGAGAGCACTACTGCAATAGAAACTTGTCTGTTATGAT-----ATTTCCTCAT 397
 QY 112 ProGlyTyrHisSerGluGluAspThrGluThrLysThrSerSerAsnGlyGluLysThr 131
 Db 398 GATGCTTACCAATTCAGAGAGATAGCGTTACCGAA-----CATCGTGCAACAGAAAGA 451

QY 132 PheLeuTyrHisSerArgAlaIleThrGluValGluTyrThrGlyLeuGlySerSerAla 151
 Db 452 TTGAGTTTCAATTCGACAGAAATGAGAGATGTCGCAAAACAGGGCTGGGCTCCGCGCA 511
 QY 152 GlyLeuValSerValAlaIleThrSerLeuSerHisPheIle-----ProAsn 168
 Db 512 GGTTCAGTACAGATTTTAACTTACAGCTTGGCTCTTTTGTATCGACCTGCAAAAT 571
 QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysThrAla 188
 Db 572 AATGTAGCAAAATATAGAGAGATTTCAATTAATTTAGCAAGTTGCTCATTTGCACCT 631
 QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
 Db 632 CAGGTAATAATTGAGAGGGGTTTGTATGATTCACGGCGAGCATATGATCTATCAGATAT 691
 QY 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGlu 228
 Db 692 AGAAGATTCACACCCGCTTATCTTAATTTGCGCAGATAT-----GGAAGTGTCT 742
 QY 229 LysPheProThrGluLeuLysLeuIle---GluSerAsnTyrGluLysHisGlu 247
 Db 743 ACTTACGGCAGTAATAGCGCCATTTGGTTGATGAGAGACTGATATTAACATTTAA 802
 QY 248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlySerGlu 267
 Db 803 AGTAACCATTTACCTTCGGGATTAACCTTTATGATAGAGCGGATTTAAGAAATGTTCA 862
 QY 268 ThrProLysLeuValSerArgValLeuGlnTyrLysGlyLysProGluGluSerSer 287
 Db 863 ACAGTAATACTGGTCCAGAGAGTAATAAATTGTATGATTCGATATCCGAAAGCTTG 922
 QY 288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
 Db 923 AATATATATACAGAACTGATCATGCAATTCGAAATTTATGATGATGATCTTAACTA 982
 QY 305 ArgGluLeuArgGlyLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHis 324
 Db 983 GATGCTTACGAGAGACTCATGATGATTAACGAGATGATATTTAGTCTCTTGAGAGG 1042
 QY 325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
 Db 1043 AATGACTGTACTGTCAAAAGTATCTGAATATCAGAAAGTATGAGATGACAGTTGCCA 1102
 QY 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspVal 355
 Db 1103 ATTGACGTTCTCTTAAAGAAATTAACCTAAGAAATCTGTGCGCATATCGAACCTCCG 1162
 QY 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyLysValValPro 375
 Db 1163 CAAACTAGCTTATTTGATGATGATGCGCAACCTTAAAGAGATTTACTTACTGCTTAAT 1222
 QY 376 GlyAlaGlyLysTyrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
 Db 1223 GTGCTGTGTGTATGACGCGCATTTGATTAAGCAAGATGTTGATCTTAAAGGCT 1282
 QY 391 GlnValGlyAsnProLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
 Db 1283 CAACCGCTAATGCAAAAG-----TTTCTTAAGGTTCAA 1318
 QY 411 TTPValAspLeuGluGlnThrGluGlyValLeuGluGlyLys---ProGluAspTyr 429
 Db 1319 TGGCTGATGTAACCTACAGCTGACGCGGTGTATAGAAAGAAAGATCCGAAACTTAT 1378
 QY 430 Ile 430
 Db 1379 CTT 1381
 RESULT 10
 US-09-918-740-60
 ; Sequence 60, Application US/0918740
 ; Publication No. US20030033626A1
 ; GENERAL INFORMATION:

OTHER INFORMATION: Operon G containing A. thaliana, S. cerevisiae, and S. pombe DNA
US-09-918-740-64

Alignment Scores:

Pred. No.:	1, 97e-82	Length:	8400
Score:	809.50	Matches:	190
Percent Similarity:	57.70%	Conservative:	76
Best Local Similarity:	41.21%	Mismatches:	146
Query Match:	36.30%	Indels:	49
DB:	11	Gaps:	14

US-10-069-062-7 (1-432) x US-09-918-740-64 (1-8400)

```

QY 3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyTyrLeuValLeuGluPro 22
DB 32 AGAGCCTTACAGTCCCGAGGAAAGCTTACTAGCGTGGAGATTTAGTTAGTACA 91
QY 23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys 42
DB 92 AAATATGAAGCAATTTGTAGTCCGATATCGCAAGATCATCTGTACCCATCCTTAC 151
QY 43 GlyThrSerLeuLeuGluSer-----ArgIleLysIleSerSerProGluPheAla 59
DB 152 GGT---TCATTCGACAGGCTTCATTAAGTGGCGTGAAAGTAAACAAATTAA 208
QY 60 AangIyGluTyrGluTyrHisIleSerSerAsnThrGlu---LysProArgGluValGln 78
DB 209 GATGGGAGTGGCTGTCATTAATTAAGTCTTAAAGCGCTTACCTCTTCGATAGGC 268
QY 79 SerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnPro 98
DB 269 GGATCTTAAGACCTTTCATTTGAAAGTATCGCTAACGCTTATTAACCT 328
QY 99 Thr-----GluAlaPheAspLeuGluIleIleIleTyrSerAsp 111
DB 329 AACATGAGACGACTCTGCAATGAAGCTTGTCTTATTAAT---ATTTCCTGAT 382
QY 112 ProGlyTyrHisSerGlnGluAspThrGluThrIleThrSerSerAsnGlyLysLysThr 131
DB 389 GATGCTTACATTTCTAGAGAGATAGCGTTACCGAA-----CATCGTGGCAACAGAGA 436
QY 132 PheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAla 151
DB 437 TTGAGTTTTCATTTGCGACAGATTTGAAGATTTCCCAAAACGGCGCTGCTCGGCA 496
QY 152 GlyLeuValSerValValAlaThrSerLeuSerHisPheIle-----ProAsn 168
DB 497 GGTTAGTACAGTTTAACTACAGCTTGGCTCTTTTGTATCGGACCTGGAAT 556
QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
DB 557 AATGTAGCAATATATAGAGAGATTTATCTAATTTAGCAACAGTTCATTTGCTACACT 616
QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
DB 617 CAGGGCTAAATTTGAAGCGGCTTATGATGAGCGGCGCATATGATGATCTATCATGAT 676
QY 209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspGlu 228
DB 677 AGAAGATTTCCACCCGCAATTAATCTTAATTTGCCAGATATT-----GGAAGTCT 727
QY 229 LysPheProThrGluLeuLysLysLeuIle---GluSerAsnThrGluLysHisGlu 247
DB 728 ACTTAGCGGAGTAATCGCGCATTTGTGTATGAGAAGACTGGAATTTACGATTAA 787
QY 248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlyLysSerGlu 267
DB 788 AGTAAACATTTACCTTCGGGATTAATCTTATGATGAGCGGCAATTAAGCATGTTCA 847
QY 268 ThrProLysLeuValSerArgValLeuGlnTyrLysGluLysProGluGluSerSer 287
DB 848 ACAGTAAACCTGCTCAGAAAGGTAATAAATTTGTATGATTCGATATGCGCATATGCGA 907

```

```

QY 288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
DB 908 AAATATATATACAGAACTGATCATCAATTTCTATTTATGAGTACATCTTAACATA 967
QY 305 ArgGluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHis 324
DB 968 GATTCCTTACACAGACATCATGACGATTAACAGCGATCATATTTAGATCTCTTGAGAG 1027
QY 325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
DB 1028 AATGACTGATCTCTCAAAAGATATCTGAAATACACAGAAATTAGAGATGCGCTCCACA 1087
QY 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspVal 355
DB 1088 ATTACACGTTCTCTTAGAATAAATATTAAGATTTGATGGCGATATGAACTCCCGTA 1147
QY 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
DB 1148 CAATCTACCTTAATGATGATGATGCGGACCTTAAAGAGATTTCTTACTGTTAATACCT 1207
QY 376 GlyAlaGlyGlyTyrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
DB 1208 GGTGCTGATGTTATGACGCCATTTGCACTGATTAAGCAAGATGTTGATTCGAGCT 1267
QY 391 GlnValGlyAsnPheLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
DB 1268 CAATCCGCTAATGACAAAAG-----TTTTCTAAGTTCA 1303
QY 411 TrpValAspLeuGluGlnThrGluGlyValLeuGluGluLys---ProGluAspTyr 429
DB 1304 TGGCTGATGTAACTACAGCTGACTGAGGTGTGTGAAAGAAAGATCCGAAACTTAT 1363
QY 430 Ile 430
DB 1364 CTT 1366

```

RESULT 12

US-10-069-909-7

Sequence 7, Application US/10006909

Publication No. US20030148479A1

GENERAL INFORMATION:

APPLICANT: KEASLING, JAY

APPLICANT: MARTIN, VINCENT

APPLICANT: PITERA, DOUGLAS

APPLICANT: KIM, SEON-MON

APPLICANT: WITHERS III, SYDOR T.

APPLICANT: YOSHIKUNI, YASUO

APPLICANT: NEWMAN, JACK

APPLICANT: KHEBNIKOV, ARTEM VALENTINOVICH

TITLE OF INVENTION: BIOSYNTHESIS OF ISOPENTENYL PYROPHOSPHATE

FILE REFERENCE: 2000-0007

CURRENT APPLICATION NUMBER: US/10/006, 909

CURRENT FILING DATE: 2002-04-02

NUMBER OF SEQ ID NOS: 13

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 7

LENGTH: 9253

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Synthetic

US-10-069-909-7

Alignment Scores:

Pred. No.:	2, 23e-82	Length:	9253
Score:	809.50	Matches:	190
Percent Similarity:	57.70%	Conservative:	76
Best Local Similarity:	41.21%	Mismatches:	146
Query Match:	36.30%	Indels:	49
DB:	12	Gaps:	14

US-10-069-062-7 (1-432) x US-10-006-909-7 (1-9253)

```

QY      3  LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyTYrLeuValLeuGluPro 22
      6263  AGAGCTTCAGTCCCGCCAGGAGGAGGCTTACTAGCTGGTGGATTTAGTTAGTATCA 6322
QY      23  IleTYrAspAlaTYrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys 42
      6323  AATATGAAGCATTTAGTCGATTTATCGGCAAGATGATGCTGTAGCCATCTTAC 6382
QY      43  GlyThrSerLeuLysGluSer-----ArgIleLysIleSerSerProGlnPheAla 59
      6383  GGT---TCATTCGAAAGGCTGTGATTAAGTGAAGTCCGTGTAAGAAACAAATTAA 6439
QY      60  AsnGlyLysIleProGlyLysIleSerSerSerAlaGlu---LysProArgGluValGln 78
      6440  GATGGGAGTGGCTGATCACTATTAAGTCTTAAAGGCTTACTTCTGTTGATAGGC 6499
QY      79  SerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTYrIleGlnPro 98
      6500  GGATCTAAGAACCTTTCAATTGAAAAAGTATCGTACAGTATTTAGTAAACCT 6559
QY      99  Thr-----GluAlaPheAspLeuGluIleIleIleTYrSerAsp 111
      6560  AACATGACGACTACTGCAATAGAACTTGTCTGTTATGAT-----ATTCTCTGAT 6613
QY      112  ProGlyTYrHisSerGlnGluAspThrGluThrLysThrSerSerAsnGlyLysLysThr 131
      6614  GATGCTTACCTTCTCAGAGAGATAGCTTACCGAA-----CATGTTGGCAACAGAA 6667
QY      132  PheLeuTYrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAla 151
      6668  TTGAGTTTCATTCGCAAGAAATTGAAGAAATCCCAAAAAGGGCTGGCTCTCGGCA 6727
QY      152  GlyLeuValSerValValAlaThrSerLeuLeuSerHisPheIle-----ProAsn 168
      6728  GGTTTAGTACACATTTTAACTACAGCTTGGCTCTTTTGTGATCGGACCTGGA 6787
QY      169  ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleHisCysTYrAla 188
      6788  AATGTAGCAAAATATAGAGAGATTTATTCATATTGACAGATGCTCTCATCTCAAG 6847
QY      189  GlnLysLysIleGlySerGlyPheAspValAlaIleTYrGlyLeuIleValTYr 208
      6848  CAGGGTAAATTTGAGAGCGGTTGATGAGCGGCGAGCATATGATCTATCAATAT 6907
QY      209  ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGlu 228
      6908  AGAAGATTCACACCCGATTAATCTTAATTGCGAGATAT-----GGAGTGTCT 6958
QY      229  LysPheProThrGluLeuLysLysLeuIle---GluSerAsnProGluLysHisGlu 247
      6959  ACTTACGGCAGTAATCTGGCGCATTTGTTGATGAGAAAGACGCAATATTCAGATTAA 7018
QY      248  ArgCysThrLeuProTYrGlyIleLysLeuLeuMetGlyAspValLysGlySerGlu 267
      7019  AGTAACCATTTTACCTTCGGGATTAATCTTATGATGGGCGATTAAGATGGTTCAGA 7078
QY      268  ThrProLysLeuValSerArgValLeuGlnTrpLysGluLysProGluGluSerSer 287
      7079  ACGTAAACCTGTCAGAAAGTAAAAATTTGATGATTCGATATGCCAGAAACCTTGG 7138
QY      288  ValIleValAspGlnLeuLysSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
      7139  AAAATATATACAGAACTCATCATGCAATTTCAAGTTTATGATGAGCATATCACTA 7198
QY      305  ArgGluMetArgGluLysTYrAspSerAspProGluThrTYrIleLysGluLeuAspHis 324
      7199  GATCGCTTACAGAGACTCATGACATTAACAGCGCATCAATATTGAGTCTGTAGAG 7258
QY      325  Ser-----ValGlnProLeuThrValAlaIleLysAsn 335
      7259  AATGACTGATCCTGCAAAAGTATCTGAAATCAAGAAATGATGAGATCAGTTGCCA 7318

```

```

QY      336  IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspVal 355
      7319  ATTACACGTCTCTTAGAAAAATACTAAAGATCTGCGGATATGAACTCCCGTA 7378
QY      336  GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValGlyValValPro 375
      7379  CAACTAGCTTATTTGATGATGATGCGACGACCTTAAAGAGTTCCTTACTTAACTAC 7438
QY      376  GlyAlaGlyLysTYrAspAlaIleLeuValLeuAla-----LeuGluAsn 390
      7439  GGTGCTGGTGTATTAAGACGCACTTGACATGATTAACAAGATGTTGATTTAGGCT 7498
QY      391  GlnValGlyAsnPheLysGlnLysThrLeuGluAsnProAspTYrPheHisAsnValTYr 410
      7499  CAACCCGCTAATGACAAAAGA-----TTTCTAAGTTCA 7534
QY      411  TrpValAspLeuGluGlnThrGlnGlyValLeuGluGluLys---ProGluAspTYr 429
      7535  TGGCTGATGTAACTACAGCTGACCTGGGCTGTAGAAAGAAAGATCCGAAACTTAT 7594
QY      430  Ile 430
      7595  CTT 7597
QY      430  Ile 430
      7595  CTT 7597

RESULT 13
US-09-918-740-72
; Sequence 72, Application US/09918740
; Publication No. US20030033626A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; APPLICANT: Kuehnle, Adelheid
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways t
; FILE REFERENCE: KAS-103XC1
; CURRENT APPLICATION NUMBER: US/09/918,740
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221,703
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 72
; LENGTH: 13917
; TYPE: DNA
; ORGANISM: Artificial Sequence
; NAME/KEY: misc feature
; LOCATION: (1..1)
; OTHER INFORMATION: Plasmid transformation vector pHK04, containing Operon B, contain
; OTHER INFORMATION: 1
US-09-918-740-72

Alignment Scores:
Pred. No.: 4.34e-82
Score: 809.50
Percent Similarity: 57.70%
Best Local Similarity: 41.21%
Query Match: 36.30%
DB: 11 Gaps: 14

US-10-069-062-7 (1-432) x US-09-918-740-72 (1-13917)

QY      3  LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyTYrLeuValLeuGluPro 22
      4141  AGAGCTTCAGTCCCGCCAGGAGGAGGCTTACTAGCTGGTGGATTTAGTTAGTATCA 4200
QY      23  IleTYrAspAlaTYrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys 42
      4201  AATATGAAGCATTTAGTCGATTTATCGGCAAGATGATGCTGTAGCCATCTTAC 4260
QY      43  GlyThrSerLeuLysGluSer-----ArgIleLysIleSerSerProGlnPheAla 59
      4261  GGT---TCATTCGAAAGGCTGTGATTAAGTGAAGTCCGTGTAAGAAACAAATTAA 4317

```

```

QY 60 AaenGlyGluTrpGluTrpHisIIeSerSeranThru---LysProArgGluValGln 78
Db 4318 GATGGGAGATGGCTGATACATATAGTCTTAAAGAGGCTTCATCTGTTGATGAGGC 4377
QY 79 SerArgIIeAsnProPheLeuGluValaThrIlePheIleValLeuAlaTyrIleGlnPro 98
Db 4378 GGATCTAAGAACCCCTTCATGAAAAAGTTAGCTTAACGATTAAGTCTTAACCT 4437
QY 99 Thr-----GluAlaPheAspLeuGluIleIleIleTyrSerAsp 111
Db 4438 AACATGACGACTACTGCAATAGAAAGTCTGTTATGAT-----ATTTCCTGAT 4491
QY 112 ProGlyTyrHisSerGlnGluAspThrGluThrLysThrSerSeranGlyGluTyrThr 131
Db 4492 GATGCTACCATCTCTAGAGAGTACCGTTAACCGAA-----CATGCTGCAACAGAGA 4545
QY 132 PheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerAla 151
Db 4546 TTGAGTTTCATTCGACAGAAATGAGAAATGCCAAAACAGGCTGGCTCTCGGCA 4605
QY 152 GlyLeuValSerValValaThrSerLeuLeuSerHisPheIle-----ProAsn 168
Db 4606 GGTTCATGACAGTTTAACTACAGCTTGCTCTTTTGTGATCGACCTCGAAAT 4665
QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValaIleGlnIleAlaHisCysTyrAla 188
Db 4666 AATGTGACAAATATATAGAGATTTATTCATTAATTGACCAAGTCTCATTTGCAAGCT 4725
QY 189 GlnLysLysIleGlySerGlyPheAspValaIleThrAlaIleTyrGlyLeuIleValTyr 208
Db 4726 CAGGGTAAATTTGAGAGCGGGTTGATGAGCGCGGAGCATATGATCTATCAGATAT 4785
QY 209 ArgArgPheGlnProAlaLeuIleAsnAspValaPheGlnValLeuGluSerAspProGlu 228
Db 4786 AAGAGTTCCACCGCATTAATCTTAATTTGCCAGATAT-----GAAAGTGT 4836
QY 229 LysPheProThrGluLeuLysLysLeuIle---GluSerAsnTrpGluLysLysGlu 247
Db 4837 ACTTACGCGAGTAACTGGCGCATTTGTTGATGAGAGACTGGAATTTACGATTA 4896
QY 248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlySerGlu 267
Db 4897 AGTAAACATTTACCTTCGGGATTAATCTTAATGATGGGAGATTAAGATGTTACAGAA 4956
QY 268 ThrProLysLeuValSerArgValLeuGlnTyrLysGluLysProGluGluSerSer 287
Db 4957 AAGGTAACCTGCTCCAGAAAGTAAAAAATGATGATGATCCCATATGCCAGAAAGCTTG 5016
QY 288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGluLeu 304
Db 5017 AAAATATATACAGAACTGATCATGCAATTTAGATTTATGAGATGATCTTAACCT 5076
QY 305 ArgGluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAspHis 324
Db 5077 GATCGCTTACAGAGACTGATGACATTAACAGCATATGATTTGAGCTCTTGAGAGG 5136
QY 325 Ser-----ValGluProLeuThrValaIleLysAsn 335
Db 5137 AATGACTGACTCTGCAAAAATGATCTGTAATCAAGAAATGAGATGATCCACCA 5196
QY 336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGluValaProIleGluProAspVal 355
Db 5197 ATTACGCTTCTTAAAGAAATTAAGAAATCGGCGGAGATGAGATCCCTCCGTA 5256
QY 356 GlnThrGlnLeuLeuAspArgCysGlnGluIleProGlyCysValaGlyValaValaPro 375
Db 5257 CAATCTAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5316
QY 376 GlyAlaGlyGlyTyrAspAlaIleAlaValLeuVal-----LeuGluLysn 390
Db 5317 GGTGCTGCTGATTAAGACCCATGATGATGATGATGATGATGATGATGATGATGAT 5376
QY 391 GlnValaGlyAsnPheLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410

```

```

Db 5377 CAACCGCTTAATGACAAAAGA-----TTTCTAAGGTTCA 5412
QY 411 TrpValaPheLeuGluGlnGlnThrGluValaLeuGluLys---ProGluAspTyr 429
Db 5413 TGGCTGATGATTAAGTACGAGCTGATGAGGCTGTTAGAGAAAGAAAGATCGAAACTAT 5472
QY 430 Ile 430
Db 5473 CTT 5475

RESULT 14
US-09-918-740-74
; Sequence 74, Application US/09918740
; Publication No. US20030033626A1
GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; APPLICANT: Kuehnle, Adelheid
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
; FILE REFERENCE: KAS-103XC1
CURRENT FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: 60/221,703
PRIOR FILING DATE: 2000-07-31
NUMBER OF SEQ ID NOS: 76
SOFTWARE: PatentIn version 3.0
SEQ ID NO 74
LENGTH: 14623
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: misc feature
LOCATION: (1..7)
OTHER INFORMATION: Plastic transformation vector pKH08, containing Operon C, contain
; OTHER INFORMATION: 1
US-09-918-740-74

Alignment Scores:
Pred. No.: 4,696-82 Length: 14623
Score: 809.50 Matches: 190
Percent Similarity: 57.70% Conservative: 76
Best Local Similarity: 41.21% Mismatches: 146
Query Match: 36.30% Indels: 49
DB: 11 Gaps: 14

US-10-069-062-7 (1-432) x US-09-918-740-74 (1-14623)
QY 3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyTyrTyrLeuValLeuGluPro 22
Db 2625 AGAGCTTCAAGTCCCGCCAGGAAAGCTTACTAGCTGATGATTTAGTTAGATAC 2684
QY 23 IleTyrAspAlaTyrValaThrAlaLeuSerSerArgMetHisAlaValaIleThrProLys 42
Db 2685 AAAATATGACATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2744
QY 43 GlyThrSerLeuLysGluSer-----ArgIleLysIleSerSerProGlnPheAla 59
Db 2745 GGT---TCATTTGACAGGCTGATGATGATGATGATGATGATGATGATGATGAT 2801
QY 60 AaenGlyGluTrpGluTyrHisIIeSerSeranThru---LysProArgGluValGln 78
Db 2802 GATGGGAGATGGCTGATACATTAAGTCTTAAAGAGTGGCTTCATCTGTTGATGAGC 2861
QY 79 SerArgIleAsnProPheLeuGluValaThrIlePheIleValLeuAlaTyrIleGlnPro 98
Db 2862 GGATCTAAGAACCCCTTCATGAAAAAGTTATCGCTAAGATTAAGTACTTAACCT 2921
QY 99 Thr-----GluAlaPheAspLeuGluIleIleIleTyrSerAsp 111
Db 2922 AACATGACGACTACTGCAATAGAAAGTCTGTTATGAT-----ATTTCCTGAT 2975
QY 112 ProGlyTyrHisSerGlnGluAspThrGluThrLysThrSerSeranGlyGluTyrThr 131

```

```

Db      2976 GATGCTACATTCCTCAGGAGGATAGCGTTACCGAA-----CATGCTGGCAACAGAAAGA 3029
Qy      132 PheLeuTYrHisSerArgAlaIleThrGluValGluLeuThyGlyLeuGlySerSerAla 151
Db      3030 TTGAGTTTCATTCCTCGACAGAAATGGAAGATTCACCAAAACAGGGCTGGCTCTCCGGCA 3089
Qy      152 GlyLeuValSerValValAlaThrSerLeuLeuSerHisPheIle-----ProAn 168
Db      3090 GGTTCATGACAGATTTTAATACAGCTTGGCTCTCTTTTGTATCGAGACCTGGAAAT 3149
Qy      169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
Db      3150 AATGTGACAAATATATAGAGAGATTAATTCATTAATTGACACAGAGTCTCATGTCAAGCT 3209
Qy      189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
Db      3210 CAGGGTAAATGGAACGGGGTTGATGAGCGCGCGACATATGATCTATCAGATAT 3269
Qy      209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGlnSerAspProGlu 228
Db      3270 AGAAGATTCACACCCCATTAATCTTAATTTGCCAGATAT-----GGAAGTGCCT 3320
Qy      229 LysPheProThrGluLeuLysLysLeuIle---GluSerAsnTyrGlnLysLysGlu 247
Db      3321 ACTTAGCGAGTAATCTGGCGCATTTGGTTGATGAAGAAAGACGTGAATATACGATTAA 3380
Qy      248 ArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlySerGlu 267
Db      3381 AGTAACCATTTTACCTTCGGGATTAATCTTAATGATGGGCGATATTAAGATGGTTACGA 3440
Qy      268 ThrProLysLeuValSerArgValLeuGlnTyrLysGlnLysProGlnLysSer 287
Db      3441 ACGATAAACTGCTCCAGAAAGTAAATAATGCTGATGATTCGATATGCCAGAAACTTG 3500
Qy      288 ValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMet-----LysGlnLeu 304
Db      3501 AAAATATATATACAGAACTCATCATCATCAAAATTCATGAATTAATGATGACATCTAAACA 3560
Qy      305 ArgGluMetArgGluLysTyrAspSerAspProGlnTyrTyrIleLysGlnLeuAspHis 324
Db      3561 GATCGCTTACACGAGACTCATGACATTAACGAGATCAGATATTAAGTCTTGAAGAG 3620
Qy      325 Ser-----ValGluProLeuThrValAlaIleLysAsn 335
Db      3621 AATGACTGATACCTGTCAAAAGTATCCTGAATATCAGAAATGAAGATGACAGTGGCA 3680
Qy      336 IleArgLysGlyLeuGlnAlaLeuThrGlnLysSerGlnValProIleGluProAspVal 355
Db      3681 ATTAGACGTTCTTTAGAAAATAATTAATAAGATCTGGTCCGATATCGAAGCTCCGTA 3740
Qy      356 GlnThrGlnLeuLeuAspArgCysGlnGlnIleProGlyCysValGlyValValPro 375
Db      3741 CAATACGCTTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3800
Qy      376 GlyValGlyValTyrAspAlaIleAlaValLeuVal-----LeuGluAsn 390
Db      3801 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3860
Qy      391 GlnValGlyAsnLysLysGlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyr 410
Db      3861 CAATACGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3896
Qy      411 ThrValAspLeuGlnLysGlnThrGlnLysValLeuGlnLys---ProGluAspTyr 429
Db      3897 TGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3956
Qy      430 Ile 430
Db      3957 CTT 3959

```

RESULT 15
US-09-918-740-76

```

/ Sequence 76, Application US/09918740
/ Publication No. US20030033626A1
/ GENERAL INFORMATION:
/ APPLICANT: Hahn, Frederick
/ APPLICANT: Kuehnle, Melheid
/ TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
/ TITLE OF INVENTION: create novel traits in transgenic organisms
/ FILE REFERENCE: KAS-103XC1
/ CURRENT APPLICATION NUMBER: US/09/918,740
/ PRIOR FILING DATE: 2001-07-31
/ PRIOR APPLICATION NUMBER: 60/221,703
/ PRIOR FILING DATE: 2000-07-31
/ NUMBER OF SEQ ID NOS: 76
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 76
/ LENGTH: 14623
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1..7)
/ OTHER INFORMATION: Plasmid transformation vector pRHO6, containing Operon E, contain
/ US-09-918-740-76

Alignment Scores:
Pred. No.: 4,696-82 Length: 14623
Score: 809.50 Matches: 190
Percent Similarity: 57.70% Conservative: 76
Best Local Similarity: 41.21% Mismatches: 146
Query Match: 36.30% Indels: 49
DB: 11 Gaps: 14

US-10-069-062-7 (1-432) x US-09-918-740-76 (1-14623)

Qy      3 LysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyGlyTyrLeuValLeuGluPro 22
Db      2625 AGAGCCTTCAGAGCCCGGAGAAACGTTACTAGCTGGTGGATGATTTGATTAGATACA 2684
Qy      23 IleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLys 42
Db      2685 AATATATGAAGCATTTGTGATCGGATTAATCGCAAGAAATGATGATGATGATGATGATGAT 2744
Qy      43 GlyThrSerLeuLysGluSer-----ArgIleLysSerSerProGlnPheAla 59
Db      2745 GGT---TCATTCGACAGGCTCTGATTAAGTTGAAGCGCGTGAAGAAAGTAAACATTTAA 2801
Qy      60 AsnGlyLysTyrGlyLysTyrHisLeuSerSerAsnThrGlu---LysProArgGluValGln 78
Db      2802 GATGGGGAGTGGCTGTACCATTAATTAAGTCTTAAGTCTCATTCCTGTTTCGATAGGC 2861
Qy      79 SerArgIleAsnProPheLeuGlnAlaThrIlePheIleValLeuAlaTyrIleGlnPro 98
Db      2862 GATCTTAAGAACCCCTTTCATTTGAAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 2921
Qy      99 Thr-----GluAlaPheAspLeuGlnIleIleTyrSerAsp 111
Db      2922 AACATGACAGACTACTGCAATGAATAACTGTTCTGTTATTTGAT-----ATTTCCTGAT 2975
Qy      112 ProGlyTyrHisSerGlnGlnAspThrLysThrLysSerSerAsnGlyLysThr 131
Db      2976 GATGCTACATTCCTCAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3029
Qy      132 PheLeuTYrHisSerArgAlaIleThrGluValGluLeuThyGlyLeuGlySerSerAla 151
Db      3030 TTGAGTTTCATTCCTCGACAGAAATGGAAGATTCACCAAAACAGGGCTGGCTCTCCGGCA 3089
Qy      152 GlyLeuValSerValValAlaThrSerLeuLeuSerHisPheIle-----ProAn 168
Db      3090 GGTTCATGACAGTTTTAATACAGCTTGGCTCTCTTTTGTATCGAGACCTGGAAAT 3149
Qy      169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188

```



```

Db 730 GCTGCCATTACGGCTCTGTCTACAGACGCTTCTCCCTCGATTCTCGAATCAGTG 789
Qy 220 -----PheGluValLeuGluSerAspPro 227
Db 790 GGTGACGCTGTTCTCCAGGCTTCCAGAGCGGTTGTTCCGAGTGTAAAGAGACGCCGAC 849
Qy 228 GluYrPheProThrGluLeuYrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 247
Db 850 CCCGACGATCCG-----TGGGACACCGAA----- 873
Qy 248 ArgCys-----ThrLeuProTyrGlyLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 262
Db 874 ---TCTTACAGCTTCGCGATGAGCTCCCGCGAATGCAATGCTCTCTGTGATCGTC 930
Qy 263 LysGlyGlySerGluThrProLysLeuValSerArgValLeuGluThrLysLeuGlyLys 282
Db 931 GAAATGTTGTTCCAGACTTCATCCATGAGTGAAGAAAGTTTGAGTGGGAGGAAACAGAT 990
Qy 283 ProGluGluSerSerValValTyrAspGluLeuAsnSerAlaAsnLeuGlnPheMetLys 302
Db 991 CAGAAAGAACCGCATATGCTCTGGGGCGCTCTGCAATGAAACAGAGACTTCGCCGTC 1050
Qy 303 GluLeuArgGluMetArgGluLysTyrAspSerAspProGluThrTyrLeuGluLeu 322
Db 1051 GAACTCAGACGCTTGACAGAGCCCGGAC----- 1080
Qy 323 AspHisSerVal-----GluProLeuThrValAlaLeuLysAsnLeuArgLysGly 339
Db 1081 GAACTACTCTCAGTGCATTTGAAATGTCGCGACCTTATTCACGCTCGCGTAACAC 1140
Qy 340 LeuGlnAlaLeuThrGluLysSerGluValProLeuGluProAspValGlnThrGluLeu 359
Db 1141 ATCCGTTCCATGACTCAAAAGTCGATGCCAATCGAGCCGCGGTCCAAACGAGCTA 1200
Qy 360 LeuAspArgCysGlnGluLeuProGlyCysValGlyValValProGlyValGlyGly 379
Db 1201 CTGACGCTCTCTCGAGTGGAGGTGTCATCGGTGTGTGTTCCAGAGCAAGGAGGAC 1260
Qy 380 TyrAspAlaLeuValLeuValLeuGluAsnGln-----ValGlyAsnPheLysGln 397
Db 1261 TACGAGCGCATGTGCTCTCTCATCCAAAGCAATCCGAGTGTATCACCAGATTGAAACC 1320
Qy 398 -----LysThrLeuGluAsn-----ProAspTyr-----PheHisAsnVal 409
Db 1321 TTCCTTGAGACTTGGGAGAGCAAGCGAGACGATTTCCGTGGCAAGATTGGGAAATC 1380
Qy 410 TyrTrpValAspLeuGluGluGlnThrGluGlyValLeuGluGluLysProGluAspTyr 429
Db 1381 AGGCTTCTTGGCGTCCGCCATGATCAGAGGAGTCAAGAACGAGATGCTCGAGCATAT 1440
Qy 430 IleGly 431
Db 1441 GCGGGC 1446

```

```

; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1406
; LENGTH: 1509
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-1406

Alignment Scores:
Pred. No.: 4,76e-65 Length: 1509
Score: 649.50 Matches: 166
Percent Similarity: 47.88% Conservative: 83
Best Local Similarity: 31.92% Mismatches: 150
Query Match: 29.13% Indels: 121
DB: 14 Gaps: 16

US-10-069-062-7 (1-432) x US-10-128-714-1406 (1-1509)
Qy 4 AlaPheSerAlaProGlyLysValPheLeuAlaGlyLysThrLeuValLeuGluProIle 23
Db 28 GCTTTGTCGCGCGCGGAGAGTCTCTCTACTGGGGTTATCTGTGCTGAGACCGCAGC 87
Qy 24 TyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIle----- 39
Db 88 TACACTGGAGTGTGTTGCTTGGCCCTTGAGCGCAGAAATCCATGTCATGTTCCAGCAATTGAGA 147
Qy 40 -----ThrProLys 42
Db 148 CGGAACATCCGCGGAGACCGCCTCGGATCCGCGATGCCGTCGACACACTCAG 207
Qy 43 -----GlyThrSerLeuGluSerArgIleLysIleSerSerPro 56
Db 208 GCGAAGGCAATGTTTCATGAGACAGACAGACAGACAGATGCTGTACCTCCCA 267
Qy 57 GlnPheAlaAsnGlyGluTyrGluTyrHisIle----- 67
Db 268 CAGTTTCGTGATGGGATGAGATGATATGATACAAAGATCGAGAGCGTGAAGATC 327
Qy 67 ----- 67
Db 328 CTAGTGAACAGAGAAATGATGGTAAAGCAATATTATATGATCAGTAATTTGCCCA 387
Qy 68 SerSerAsnThrGluLysProArgGluValGlnSerArgIleAsnProPheLeuGluAla 87
Db 388 GAGCTGACCATCAGGCGACGC-----AACCTGTTGTGAGACCC 429
Qy 88 ThrIlePheLeuValLeuAlaTyrIle-----GlnProThrGluAlaPheAsp 103
Db 430 TCTCTGAACCTTGGCTTGACTTACATCAGATATGCGGCACTGGAAGATTGGGGTCA 489
Qy 104 LeuGluLeuIleIleTyrSerAspProGlyTyrHisSerGlnGluAspThrGluThrLys 123
Db 490 TTATCGATTACTATCTCCCGCAGACAGATTAATCTCCGAGAGCGCTTCTCCAGAGCT 549
Qy 124 ThrSerSerAsnGlyLysThrPheLeuYrHisSerArgAlaIleThrGluValGlu 143
Db 550 TCGGACTCCGCTGTCAGAGATTCGTGACTTGTGCTTGCCTTCAGAGGACAC 609
Qy 144 LysThrGlyLeuGlySerSerAlaGlyLeuValSerValAlaIleThrSerLeuLeuSer 163
Db 610 AAGACAGCCCTGGCTTCAAGCGCTTGTATCAGCTGCTGCTGTCTGTATC 669
Qy 164 His-----PheIleProAsnValIleSerThrAsnLysAspIleLeuHisAsnValAla 181
Db 670 CACCGTACTATGCAACCCGACGATCTCGTCCAGGCGCGACGACGCTTCAACATCTGCGC 729

```

```

QY 182 GlnlleAlahisCystrAlaInuylsyleGlySerGlyPheAspValAlaThrAla 201
   |||||
Db 730 CAGGGGCCCCACGCGCGTGCAGGGTAAAGTCGGGTCCGGCTTCATGTGGAGCTGCC 789
QY 202 lletYrGlyLeuileValYrArGArpGheGlnProAlaLeuileAsnAspVal 219
   |||||
Db 790 ATTTCAGGCTCCTGCTCTCAAGACCGCTTCTCCCTCCATCTCCAAATCAGTGGGTGAC 849
QY 220 -----PheGlnValLeuGlnSerAspProGluYls 229
   |||||
Db 850 GCTGCTTCCAGGCTTCCAGAGCGGTTGTCGGATGCTAGACGCCGCCAGCCCCAG 909
QY 230 PheProThrGlnLeuYlsyleLeuileGlnSerAspTrpGlnYlsyleGluArgCys 249
   |||||
Db 910 CATCCG-----TGGGACCCGAA-----TGT 930
QY 250 -----ThrLeuProYrGlylleYlsyleLeuLeuMetGlyAspValYlsGly 264
   |||||
Db 931 CTAGACTTCGCGATGAGCTCCCGCGAATGCAATGGTCTCTGTGATCGTCAATGT 990
QY 265 GlySerGlnThrProYlsLeuValSerArGValLeuGlnTrpYlsyleGluYlsProGlu 284
   |||||
Db 991 GGTTCGACATCTCCATCCATGTTGAGAAAGTTTGGAGTGGCGGAAACGAAATCAGAG 1050
QY 285 GluSerSerValYrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetYlsGlnLeu 304
   |||||
Db 1051 GAAGCCGATATGCTCTGGGGCGCTCTGCAATGCAACAGAGACTTGGCCCTGGAATC 1110
QY 305 ArgGluMetArGlnYlsYrAspSerAspProGluThrYrileYlsGlnLeuAspHis 324
   |||||
Db 1111 AGACGCTTGACAGAGCCCGGAC-----GAACTAT 1140
QY 325 SerVal-----GluProLeuThrValAlaileYlsAsnilleArgYlsGlnLeuGln 341
   |||||
Db 1141 ACTCTAGTGACTTTGAAATATGTCGACACTTATATTCGCGCTCCGCTGATCAATCCGT 1200
QY 342 AlaLeuThrGlnYlsSerGlnValProileGluProAspValGlnThrGlnLeuLeuAsp 361
   |||||
Db 1201 TCCATGATCTCAAAAGTCGATGTCCTCAATCGAGCCGCGGTCAAAACGAGTACTTGAAC 1260
QY 362 ArgCysGlnGlnileProGlyCysValGlyValValProGlyValaglyYrAsp 381
   |||||
Db 1261 GCTCTGTCGAGTGGAGGTGTCATCGGTGTGTTCGAGACAGGGGCTTACGAC 1320
QY 382 AlaileAlaValleuValLeuGlnAsnGln-----ValGlyAsnPheYlsGln----- 397
   |||||
Db 1321 GCCATGTGCTCTCTCAATCCAGAACATCCGATGTGATCCAGATTGAAAGCTTCTTT 1380
QY 398 LysThrLeuGlnAsn-----ProAspYr-----PheHisAsnValYrTrp 411
   |||||
Db 1381 GAGACTTGGAGAGCAAAAGCGAGACGATTTCCGTGGCAAAATTTGGAAAGTCAGCGCT 1440
QY 412 ValAspLeuGlnGlnGlnThrGlnGlyValLeuGlnGlnYlsProGluAspYrileGly 431
   |||||
Db 1441 CTTGGGCTCGCGCATGATCAGAGGATGCAAGAACGAGATGCTCGAGCAATATGCGGCG 1500

RESULT 19
US-10-128-714-6406
; Sequence 6406, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Mengli
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroszhkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128, 714
; PRIOR FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697

```

```

; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6406
; LENGTH: 1509
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-6406

Alignment Scores:
Pred. No.: 4,766-65 Length: 1509
Score: 649.50 Matches: 166
Percent Similarity: 47.88% Conservative: 83
Best Local Similarity: 31.92% Mismatches: 150
Query Match: 29.13% Indels: 121
DB: 14 Gaps: 16

US-10-069-062-7 (1-432) x US-10-128-714-6406 (1-1509)

QY 4 AlaPheSerAlaProGlyYlsAlaPheLeuAlaGlyYrTrpLeuValLeuGlnProile 23
   |||||
Db 28 GCTTGTCCGCGCGGGAAGGTCTCTCACTGGGGGTATCTGCTTGGACCGCAGC 87
QY 24 TyrAspAlaYrValThrAlaLeuSerSerArGMeChleAlaValile----- 39
   |||||
Db 88 TACACTGGAGCTGTGTGGCTTACCGCCAGATCCATGTCATGTCCTTACGAAATTGAGA 147
QY 40 -----ThrProYls 42
   |||||
Db 148 CGAATCATCGCGGGAAGCCGCTCGGATCCGCGATGCGCGGTGCGACACACTGAG 207
QY 43 -----GlyThrSerleuYlsGlnSerArGilleYlsileSerPro 56
   |||||
Db 208 GCGAAGCAATGTTTCATGAGACAGAAAGAGAGAGCAGATGTCATCATCTCCCA 267
QY 57 GlnPheAlaAsnGlyLutrpGluYrHisile----- 67
   |||||
Db 268 CAGTTCGTGAGTGGGTATGGAGATATGATACATCAAAAGATGCGAGACGCTGAGAGATC 327
QY 67 ----- 67
   |||||
Db 328 CTAGTGAACAGAGAAATGATGCTAAAGCAAAATATTATGATCAGTAATTGCCCA 387
QY 68 SerSerArnThrGlnYlsProArGlnValGlnSerArGilleAsnProPheLeuGlnAla 87
   |||||
Db 388 GGAGCTGACCATCCAGGCCAGC-----AACCTGTTGTGCAAGCC 429
QY 88 ThrilePheileValleAlaYrile-----GlnProThrGlnAlaPheAsp 103
   |||||
Db 430 TCTCTGAATTCGCTTGAATTCATCAGCTATGCGCGCACTCGAAGATTTCCGGTCA 489
QY 104 LeuGlnilleleleYrSerArpProGlyYrHisSerGlnGlnAspThrGlnYrlyls 123
   |||||
Db 490 TTATCGATTACTATCTCGCGCAAGATTAATCTCCGAGACGCGCTTCTCCAGGCT 549
QY 124 ThrSerSerArnGlyLutrpThrPheLeuYrHisSerArGAlaileThrGlnValGlu 143
   |||||
Db 550 TCGGAGCTCCGCTCGTCAAGACGATTCGTCGATTTGTTGCTTCAGAGACAGAC 609
QY 144 LysThrGlyLeuGlySerSerAlaGlyLeuValSerValAlaAlaThrSerleuSer 163
   |||||
Db 610 AAGACAGGCTGGGCTTTCAGCGCTTGGTCACTGCTGCTGCTCTCTCTCATC 669
QY 164 His-----PheileProAsnValileSerThrAsnYlsAspilleLeuHisAsnValAla 181
   |||||

```

Db 670 CACCGTACTATGCAACCGGAGCATCTCGGTCCAGGCGCGGACAGCACTTCACATCTGGCC 729
Qy 182 GlnIleAlaHisCysTyrAlaGlnIleValIleGlySerGlyPheAspValAlaThrAla 201
Db 730 CAGGGGGCCCACTGGGCGCGGAGGTAAGTCGGGTCCGGTTCGATGTTGCTGACCTGCC 789
Qy 202 IleTyrGlyLeuIleValTyrArgPheGlnProAlaLeuIleAsnAspVal----- 219
Db 790 ATTTCAGGCTCTGCTCTCAAGACGCTTCCTCCCTCGATTCGTAATGAGTGGGTGAC 849
Qy 220 -----PheGlnValLeuGlnSerAspProGlnIle 229
Db 850 GCTGTTCTCCAGGCTTCGAGAGCGGTTGTCGAGTCAAGAGACCGGACCCCGAG 909
Qy 230 PheProThrGlnLeuIleValLeuIleGlnSerAspProGlnIleValIleGlnIle 249
Db 910 CATCG-----TGGACACCGAA-----TGT 930
Qy 250 -----ThrLeuProTyrGlyIleValIleValLeuIleValIleValIleValIle 264
Db 931 CTAGACTTCGCGATGAGCTCCCGCGGAAATGCAATGCTCTGTCGATGATGT 990
Qy 265 GlySerGlnIleProGlyLeuValSerArgValIleGlnIleTyrValIleGlnIleVal 284
Db 991 GGTTCGACAGCTCCATCCATGCTGAGAAAGTTTGGAGTGGCGGAAACAGATCAGAG 1050
Qy 285 GluSerSerValIleValTyrAspIleValIleValIleValIleValIleValIleVal 304
Db 1051 GAGCCGATATGCTTGGGCGCTTCGCAATGCAACAGAGAGCTTCGCTGAGATC 1110
Qy 305 ArgGluMetArgGlnIleTyrAspSerAspProGlnIleTyrIleValIleValIleVal 324
Db 1111 AGAGCTTGGGACAGAGCCCGAC-----GAAAT 1140
Qy 325 SerVal-----GlnProLeuThrValAlaIleValIleValIleValIleValIleVal 341
Db 1141 ACTCGACTGACTTGAATAATGTCCGACCTATATATGACGCTCCGTAACCAATCCGT 1200
Qy 342 AlaLeuThrGlnIleValIleValIleValIleValIleValIleValIleValIleVal 361
Db 1201 TCATGATCTCAAAATCGGATGTCATCCATCCAGCCGCGCTCCAAAGCTTACTTAC 1260
Qy 362 ArgCysGlnIleProGlyCysValIleValIleValIleValIleValIleValIleVal 381
Db 1261 GCTGTGCTGAGCTGAGGCTGTCATCGGTGTGTGTTCCAGAGAGCGGGGCTACAC 1320
Qy 382 AlaIleAlaValIleValIleValIleValIleValIleValIleValIleValIleVal 397
Db 1321 GCCATTGTGCTCTTCATCCAGACATCCGATGTGATCACCAGATTTGAAAGCTTCTT 1380
Qy 398 LysThrLeuGlnAsn-----ProAspTyr-----PheIleAsnValTyrTrp 411
Db 1381 GAGACTTGGAGAGAGAAACCGAGAGCAATTCGTTGGCAAGATTGGAAAGTCAAGCTT 1440
Qy 412 ValAspLeuGlnIleValIleValIleValIleValIleValIleValIleValIleVal 431
Db 1441 CTGGCGCTCCGCAATGATCAGAGGAGTCAAGAGAGATGCTGAGACATATGCGGCT 1500

RESULT 20
; Sequence 406, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Weng
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroskin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714

; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patent version 3.1
; SEQ ID NO 406
; LENGTH: 3508
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-406

Alignment Scores:
Pred. No.: 1,78e-64 Length: 3508
Score: 649.50 Matches: 166
Percent Similarity: 47.88% Conserved: 83
Best Local Similarity: 31.92% Mismatches: 150
Query Match: 29.13% Indels: 121
DB: 14 Gaps: 16

US-10-069-062-7 (1-432) x US-10-128-714-406 (1-3508)

Qy 4 AlaPheSerAlaProGlyIleValIleValIleValIleValIleValIleValIleVal 23
Db 1027 GCTTGTCCGCGCGGAGAGTCTCTCACTGGGGTTATCTGTTCTGACCGGACG 1086
Qy 24 TyrAspAlaTyrValIleValIleValIleValIleValIleValIleValIleValIleVal 39
Db 1087 TACACTGGAGCTGTGTCCTTTCAGCGCAGATTCATGCTTTCAGCAATTGAGA 1146
Qy 40 -----ThrProIle 42
Db 1147 CGGAACATCGCGGAGAGCGCTCGGATCCGCGATGCGCGGATCGGACACTGAC 1206
Qy 43 -----GlyThrSerLeuIleValIleValIleValIleValIleValIleValIleVal 56
Db 1207 GCGAGAGCAATGTTCAATGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1266
Qy 57 GlnPheAlaGlnIleValIleValIleValIleValIleValIleValIleValIleValIleVal 67
Db 1267 CAGTTCGTGATGCGGTATGAGATATGATACAAAGATGCGAGAGAGAGAGAGAGAGAG 1326
Qy 67 ----- 67
Db 1327 CTAGTAAACAGAGAAATGATGCTGTAAGCAATTTATAGTATGATGATTAATGCTTCC 1386
Qy 68 SerSerAsnThrGlnIleValIleValIleValIleValIleValIleValIleValIleVal 87
Db 1387 GAGACTACCCATCCAGCCAGC-----AACCTGTTTGTGAGAGCC 1428
Qy 88 ThrIlePheIleValIleValIleValIleValIleValIleValIleValIleValIleVal 103
Db 1429 TCTCTGAATCTCGCTTACCTTACATCATGCTATGCTGCGGACCTGGAAGATTGCGGTCA 1488
Qy 104 LeuGlnIleIleIleTyrSerAspProGlyTyrHisSerGlnIleValIleValIleValIleVal 123
Db 1489 TTATCGATTAATCTATCTCGCGAGAGAGATTAATCCAGAGCGGCTTCCAGAGCT 1548
Qy 124 ThrSerSerAsnGlnIleValIleValIleValIleValIleValIleValIleValIleVal 143
Db 1549 TCGGACTCGCGGTGTCAGAGAGATTCGTGAGACTTGTGTGCTTCCAGAGAGAGAGAG 1608
Qy 144 LysThrIleValIleValIleValIleValIleValIleValIleValIleValIleValIleVal 163
Db 1609 AAGACAGGCTTGGGCTTTCAGCGCGCTGCTGATCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1668


```

Db      1610 AAGAGAGCGCTGGGCTCTTCAGCGCCGCTTGACATGCCGCTGGTCTGCTCTGCTATC 1665
Qy      164 His-----PheIleProAsnValIleSerThrAsnLysAspIleLeuHisAsnValAla 181
Db      1670 CACCGTACTATGCAACCCGAGCATCTTCGGTCCAGGCGCCGCAACAGCTTCACAACTTCGGCC 1729
Qy      182 GlnIleAlaHisCysTyrAlaGlnLysLysIleGlySerGlyPheAspValAlaThrAla 201
Db      1730 CAGCGGCGCCCACTGCGCTGCGGAGGATAAAGTCGGGTCCGGCTTCGATGTTCAGACTGCC 1788
Qy      202 IleTyrGlyLeuIleValTyrArgArpPheGlnProAlaLeuIleAsnAspVal----- 219
Db      1790 ATTATCGGCTCCTGCTCTCTACAGAGCTTCTCCCTCGATTCTCGAATTCAGTGGGTAC 1849
Qy      220 -----PheGlnValLeuGlySerAspProGlnLys 229
Db      1850 GCTGGTTCCTCAGGCTTCGAAAGCGGTGTGTTCCGATCTGTATGAGACGCGCAACCCGAG 1909
Qy      230 PheProThrGlnLeuLysLysLeuIleGlySerAsnTrpGlnLysHisGlnArgCys 249
Db      1910 CATCCG-----TGGACACCGAA-----TGT 1930
Qy      250 -----ThrLeuProTyrGlyIleLysLeuLeuMetGlyAspAspIleGly 264
Db      1991 CTAGACCTTCGGCATGAACTCCCGCGGGATGCAAAATGCTCTCTGTGAGTGCAAAGT 1990
Qy      265 GlySerGlnThrProLysLeuValSerArgValLeuGlnIntLysLysGlnLysProGln 284
Db      1991 GGTTTCGACAGATCCCATCATGATGTGAGAAAGGTGTTGGATGGCGGAAACAAATACAGAAG 2050
Qy      285 GlySerSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGlnLeu 304
Db      2051 GAAGCGCATATGCTCTGGGGGCTCTTGAAATCGAAACAGAGACTTCGCGCTGGAACTC 2110
Qy      305 ArgGlnMetArgGlnLysTyrAspSerAspProGlnThrTyrIleLysGlnLeuAspHis 324
Db      2111 AGAGCGTTGGACACAGAGCCCGAC-----GAATAT 2140
Qy      325 SerVal-----GlnProLeuThrValAlaIleLysAsnIleArgLysGln 341
Db      2141 ACTCTAGATGACTTTGAAATGATGTCGCACCTATATTCAGGCGCTGCGTAACCAATCCGT 2200
Qy      342 AlaLeuThrGlnLysSerGlnValProIleGlnProAspValGlnThrGlnLeuLysAsp 361
Db      2201 TCATATGACTCAAAAGTCGATGCTCCATTCAGACGCGCGCTGCCAAACCGAGCTACTGAC 2260
Qy      362 ArgCysGlnGlnIleProGlyCysValGlyGlyValValProGlyAlaGlyLysTyrAsp 381
Db      2261 GCTCTGTCGAGCTGAGAGGAGTCTATCGATGATGTGTTCCAGAGACGGGGGCTACGAC 2320
Qy      382 AlaIleAlaValLeuValLeuGlnAsnGln-----ValGlyAsnPheLysGln----- 397
Db      2321 GCCATATGGCTCTTCATCCAAAGACATCCGAGTGTGATCACCAATGTAAGCCTTCTT 2380
Qy      398 LysThrIleGlnLysAsn-----ProAspTyr-----PheHisAsnValTyrTrp 411
Db      2381 GAGACTTGGGAGACCAAAAGCGAGAGCAATTCGATGCGCAAGATTGGGAAAGTCAGGCTT 2440
Qy      412 ValAspLeuGlnGlnGlnThrGlnGlyValLeuGlnGlnLysProGlnAspTyrIleGly 431
Db      2441 CTGGCGCTCGCCCATGATCAAGAGAGATCAAAAGCAGATGCTCGACCAATATGCGGGC 2500

RESULT 22
US-09-988-863A-1
; Sequence 1, Application US/09988863A
; Patent No. US20020123427A1
; GENERAL INFORMATION:
; APPLICANT: Bayer AG
; TITLE OF INVENTION: Plant phosphomevalonate kinases
; FILE REFERENCE: Le A 35 018
; CURRENT APPLICATION NUMBER: US/09/988, 863A
; CURRENT FILING DATE: 2001-11-21
; NUMBER OF SEQ ID NOS: 5

```

```

/ SOFTWARE PatentIn Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 2396
/ TYPE: DNA
/ ORGANISM: Arabidopsis thaliana
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (685)..(2199)
US-09-988-863A-1

Alignment Scores:
Pred. No.: 1,68e-44 Length: 2396
Score: 475.00 Matches: 143
Percent Similarity: 43.76% Conservative: 78
Best Local Similarity: 28.32% Mismatches: 184
Query Match: 21.30% Indels: 100
DB: 10 Gaps: 15

US-10-069-0652-7 (1-432) x US-09-988-863A-1 (1-2396)
QY 1 MetSerLysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyLysLeuValLeu 20
DB 685 ATGCGTCGTGTGCTCTGCTGCTCGGAAAGTTTGATACATCGAGGCTACCTGTGATC 744
QY 21 GluProIleTyAspAlaTyValThrAlaLeuSerSerSerMetHisAlaValIleThr 40
DB 745 GAAAGCCAAATGCAAGGCTTGTTGTTAGTACAAATGCAAGGTTTAAACGGATGTAAG 804
QY 41 ProLysGlyThrSerLeuLysGlu-----SerArgIleLysIle 53
DB 805 CCAATCAACGAAGAAGTCAAGCTGAAGCTGGAAGTGGCAGTGAAGAAATGACAGATGTCAATTA 864
QY 54 SerSerProGlnPheAlaIleAsnGlyLysIleProGlyLysIleSerSerAsn-----Thr 71
DB 865 ACATCAACCAACAGCTCTCGAGA---GAAAGCATGTATTAATCTGCATCGATCATTTGACT 924
QY 72 GluLysProAlaGlyValGlnSerArgIleAsnProPheLeuGlnAlaThrIlePheIle 91
DB 922 CTTCACTGTGTCTGCAAGTATTCAGAAACCCCTTTGTAGAGCATGGATGATACAGTAT 981
QY 92 ValLeuAlaTyIleGlnProThrGlnAlaPheAsp----- 103
DB 982 GCTAATGCTGCTCATTTGGCAACCGAAGAGCAAAAGATCATTTGCAAACTCTTA 1044
QY 104 -----LeuGlnIleIleIleTySerAspProGlyLysIleSerGlnGluAsp--- 119
DB 1042 TTGCAAGGCTTGATATACAATATTAGGCTCCAAATGACTTTTACTCATATGCGAACAG 110
QY 119 ----- 119
DB 1102 ATGAAATCGGCTGGCTTCATATGACACCAAGATGCTGGTAGTACCTTGACCCGTTTGA 116
QY 120 -----ThrGluThrLysThrSerSerAsnGlyLysIleThrPheLeuTyHisSerArg 137
DB 1162 TCATCACTCAATTCATAGCTGCGGAGTCAAAATGGT-----GCTAAT 120
QY 138 AlaIleThrGluValGlnLysThrGlyLeuGlySerSerAlaGlyLeuValSerVal 157
DB 1201 TCCAAAGCTGAAGTGCAGAAAACCTGGCTTAGTCTTCGACGACATGACAAACAGCTGTG 126
QY 158 AlaThrSerLeuLeuSerHisPheIleProAsnValIleSer----- 171
DB 1261 GTTGACGCTCTGTTA---CATTAATCTT---GGAAGTGGTGAACCTATTCATCCATGTAAA 131
QY 172 -----ThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHis 185
DB 1315 GAAGGAAGTTGGCTGTGCTGATCTAATGTTATTCATATGATATACAAACAGCTCAT 1377
QY 186 CysTyralaGlnLysIleGlySerGlyPheAspValAlaThrAlaIleTyGlyLeu 205
DB 1375 TGTCTTGACAAAGGAAGGTGCGAAGTGGTTTGAATGATAGCTGTGCTGTCTATGGAAGT 143
QY 206 IleValTyTrpArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGlnSer 225

```

```

Db      1435 CAGCGTTATGTTGCTTCTCCAGAAAGTCTTGTCAATTGCTCAGGTTGACGTAACAGGT 1494
QY      226 AspPro-----GluLysPheProThrGluLeuLysLeuLeuLeuLeuLeuLeuLeuLeu 243
Db      1495 CTGCCATTAAATGAATTTATTTGTTAATAATTTGAAAG-----GGAAATATGGAGC 1542
QY      244 GluLysHISGluLysGlyThrLeuProTyrGlyLysLeuLeuLeuLeuLeuLeuLeuLeuLeu 262
Db      1543 AATAAGAGAACTGAGTCTCTTTCATCCACAGCTGATGAATCTTCTTGGAGAACTGGA 1602
QY      263 LysGlyGlySerGluThrProLysLeuValSerArgValLeuGlnThrLysGlyLys 282
Db      1603 AGTGGTGATCTCCACACATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1662
QY      283 ProGluLysSerSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLys 302
Db      1663 CCAAGAAAGGACGAGAAAGCTGGAGAAATTTGTCAGATGCAAAATTTAGAACTGGAACCT 1722
QY      303 GluLeuArgGluMetArgLysTyrAspSerAspProGluThrTyrLeuLysGluLeu 322
Db      1723 AAGCTAAAGATGTGAGCAAAATTTAGCTAAAGACCACTGGGATTTTATCTACAGATCAT 1782
QY      323 Asp-----HisSerValGluProLeuThr 330
Db      1783 AAGTCTTGTAGTGTCTACTTCTGAAAGTGGGTGTACTATGCTACTGACCAATCAAC 1842
QY      331 ValAlaIle-----LysAsnIleArgLysGly 339
Db      1843 GAAGCATATTAATAAGAACTTTAGAGGAGAGAAAGCTATGTTGAGATGCAAAATCTT 1902
QY      340 LeuGlnAlaLeuThrGlnLysSerGluValProIleGluProAspValGlnThrGlnLeu 359
Db      1903 ATGCGTCAGATGGGTGAGGGGCTTACCTCCATGAGCTTGAATCTCAACTCAACTT 1962
QY      360 LeuAspArgCysGlnGluLeuProGlyCysValGlyGlyValValProGlyAlaGlyGly 379
Db      1963 TTGGATTTCATACATGATGCTGTAAGAGATTTCTACTTCTGCTGCTGCTGCTGCTGCTG 2022
QY      380 TyrAspAlaIleAlaValLeuValLeuGluAsnGlnValGlyAsnPheLysGlnLysThr 399
Db      2023 TTTGATGCAATATTTGCAATCACTTTAGGGGATTCGGGACCAAACTGACCCAGGATG 2082
QY      400 LeuGluAsnProAspTyrPheHisAsnValTyrTrpValAspLeuGluGlnGlnThrGlu 419
Db      2083 AGTTGG-----CACATGTTTGGCTTGTGTGAGAAAGATCCACAT 2127
QY      420 GlyValLeuGluGlu 424
Db      2128 GCGGTTTGCCTAGAA 2142

RESULT 23
US-10-036-959B-5
; Sequence 5, Application US/10036959B
; Publication No.: US20030119098A1
; GENERAL INFORMATION:
; APPLICANT: E. I. du Pont de Nemours & Company
; APPLICANT: Hallahan, David L.
; APPLICANT: Keiper-Hymko, Natalie
; TITLE OF INVENTION: Genes involved in the Biosynthesis of Isopentenyl Diphosphate in
; FILE REFERENCE: CL-1792
; CURRENT APPLICATION NUMBER: US/10/036,959B
; PRIOR FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: 60/307,637
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
; LENGTH: 1509
; TYPE: DNA
; ORGANISM: Hevea brasiliensis
US-10-036-959B-5

```

```

Alignment Scores:
Pred. No.: 2,146-41
Score: 445.50
Percent Similarity: 44.84%
Best Local Similarity: 27.38%
Query Match: 19.98%
DB: 14
Gaps: 16

US-10-069-062-7 (1-432) x US-10-036-959B-5 (1-1509)

QY      1 MetSerLysAlaPheSerAlaProGlyLysAlaPheLeuAlaGlyLysTyrLeuValLeu 20
Db      1 ATGCGTGTATGTTGCTTCTGCTCCGGTAAAGGTGTGATGATGATGATGATGATGATGATG 60
QY      21 GluProIleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThr 40
Db      61 GAAGAAGCAATGACGAGGATTTACTCTGACACAAATGCTCGATTCATGCAATTTGTAAG 120
QY      41 ProLysGlyThrSerLeuLysGlu-----SerArgLysIle 53
Db      121 CTTATTTACGATGAATCAAACTGATAGTTGGCATGGGCATGACATGATGTAATTA 180
QY      54 SerSerProGlnPheAlaAsnGlyGluTrpGlyTyrHisLeSerSerAsnThrGluLys 73
Db      181 ACATCTCCCACTGACAAAG-----GAAAGCTGTGCAAAATTTGTCACTGAAAATTTAGCT 237
QY      74 ProArgGluValGlnSerArgIle-----AsnProPheLeuGluAlaThrIlePheIle 91
Db      238 CTTCAGTGTCTCTTCAAGTGCATCAAGAAAGCAATTTGTTGAAACAGACGTGCATTT 297
QY      92 ValLeuAlaTyrIleGlnProThrGlnAlaPheAsp----- 103
Db      298 GCTGATGACGCTGCACATGCAACACTTGCACAAAGTTAAGAAAGATGCTTAAACAAGTA 357
QY      104 -----LeuGlnIleIleIleTyrSerAspProGlyTyrHisSerGlnGluAsp 119
Db      358 CTCTTGCAAGCTTGATATTTACATATTAAGTATAGTACCAATGACTTCTATCATACGGAAT 417
QY      119 ----- 119
Db      418 GAGATTGAAGCATGAGACCTCCCTTGAACACCAAGATATGGCTGACCTTCTCTTT 477
QY      120 -----ThrGluThrLysSerSerAsnGlyGlyLysThrPheLeuThrHisSer 136
Db      478 TCTCATACACCTTCAATGATAGAGAACCAATGACAA----- 516
QY      137 ArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAlaGlyLeuValSerVal 156
Db      517 AACTGCAAGCTGAGAGTATGATTAACCTGATGGGTTCATCAGCAATGACCACTGCT 576
QY      157 ValAlaIleThrSerLeuLeuSerHisPhe-----IleProAsnValIleSerThrAsnLys 174
Db      577 GTAGTGTGCTTCTTCTCATCACTTGATGATGTTGATCTTTCATCTCTGTTAAGAG 636
QY      175 -----AspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
Db      637 AAGAAATTTTCTGATCTGATTTGTTGATATTAAGCCCAAACTGCCCATTTGTATGCA 696
QY      189 GluLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyrGlyLeuIleValTyr 208
Db      697 CAAGGAAAGTGGGACGATGATTTGATGTTGATGCTTGAAGTTAAGGAGCATGATGATC 756
QY      209 ArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluLysSerAspProGlu 228
Db      757 GTGCGCTTCTCTCCAGAGTCTTCTCTGCTCAGATGCTGAGAAAGATTTCA--- 813
QY      229 LysPheProThrGluLeuLysLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLys 248
Db      814 ---TTACAGAAAGTCAATTTCAACATCTTAAGAAAGAAATGGACATGAGGACATATG 870
QY      249 CysThrLeuProTyrGlyLysLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLys 267

```


Db 871 TTTTCCTGCCACCATTTGATGAGCTGTACTAGGTGACCGAGAACTGAGGATCTTCC 930
 Qy 268 ThrProLyLeuValSerArgValLeuGlnTrpLysLeuGluLysProGluLysSer 287
 Db 931 AGCCATCATATGTGTCTCTAAAGAAATGCGAAGTGTGATCTCAGAAATCCCA 990
 Qy 288 ValValTyrAspGluLeuAsnSerAlaAsn-----LeuGlnPheMet 301
 Db 991 GAAACATGGAGAAAGTTGTCAAGGCGCAATTCAGCATTGAAAGCAATTCATATTTA 1050
 Qy 302 LysGluLeuArgGluMetArgGluLysTyrAspSer----- 313
 Db 1051 ACCAAGCTCCAGAA-----GAACTTGGAGCGCGTATTAATGTGTATGACAGTTGC 1104
 Qy 314 AspProGluThrTyrLysGluLeuAspHisSerValGluProLeuThrValAlaIle 333
 Db 1105 AGCAGAAAAAATCTCAGAGAAAGTGAATGACAGCGCACTGAAACCGAGCGAAGACGTT 1164
 Qy 334 -----LysAsnIleArgLysGlyLeuGlnAla 342
 Db 1165 GTTAAAGCATTTATGATCAAGAAATGCCATGCTTCAGATCAGAAATTCATGCGCCAG 1224
 Qy 343 LeuThrGlnLysSerGluValProLysProAspValGlnThrGlnLeuLeuAspArg 362
 Db 1225 ATGGGTGAGGCTGCGAGTGTCTCGATAGAGCTGATCAACAGACTCGACTTTTGATTTCT 1284
 Qy 363 CysGlnGluLeuProGlyCysValGlyGlyValValProGlyValGlyTyrAspAla 382
 Db 1285 ACTATGATATATGATGAGAGTCTTTGGCTGAGAGTTCTCGAGCGAGTGGTTTATGCA 1344
 Qy 383 IleAlaValLeuValLeuGluAsnGlnValGlyAsnPheLysGln-----LysThrLeu 400
 Db 1345 GTCTTCGCTGTACTAGGAGCTCTGGTACCAATGCGCAAAAGCTTGAGTTCACTC 1404
 Qy 401 GluAsnProAspTyrPheHisAsnValTyrTrpValAspLeuGluGlnThrGluGly 420
 Db 1405 -----AATGTTCTGCGCTGTGGTGAAGAGACCCCTAATGCT 1443
 Qy 421 ValLeuGluGlu 424
 Db 1444 GTTTTGTAGAA 1455
 RESULT 24
 US-09-988-863A-4
 : Sequence 4, Application US/09988863A
 : Patent No. US20020123427A1
 : GENERAL INFORMATION:
 : APPLICANT: Bayer AG
 : TITLE OF INVENTION: Plant phosphomevalonate kinases
 : FILE REFERENCE: Le A 35 018
 : CURRENT APPLICATION NUMBER: US/09/988,863A
 : CURRENT FILING DATE: 2001-11-21
 : NUMBER OF SEQ ID NOS: 5
 : SOFTWARE: PatentIn Ver. 2.1
 : SEQ ID NO 4
 : LENGTH: 728
 : TYPE: DNA
 : ORGANISM: Gossypium hirsutum
 : FEATURE:
 : NAME/KEY: unsure
 : LOCATION: 715
 : OTHER INFORMATION: n can be any nucleotide
 US-09-988-863A-4

Alignment Scores:

Pred. No.: 5,08e-17 Length: 728
 Score: 231.00 Matches: 76
 Percent Similarity: 43.63% Conservative: 37
 Best Local Similarity: 29.34% Mismatches: 60
 Query Match: 10.36% Indels: 86
 DB: 10 Gaps: 11

US-10-069-062-7 (1-432) x US-09-988-863A-4 (1-728)

Qy 17 TyrLeuValLeuGluProIleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHis 36
 Db 7 TACGCTATTTGTAAGCAATTCATCAACT----- 36
 Qy 37 AlaValIleThrProLysGlyThrSerLeuLysGluSerArgIleLysIleSerPro 56
 Db 37 -----ATCAAGCCGTAAGAGCTGGGATGCTTGTCCAGATGTCATCACTCTCT 90
 Qy 57 GlnPheAlaAsnGlyLysTrpGluTyrHisIleSer----- 68
 Db 91 CAGCTTCCAG-----GAAAGCATGTATAAATGTTCTCGAAACATTAACTTCACTGCT 147
 Qy 69 ---SerAsnThrGluLysProArgGluValGlnSerArgIleAsnProPheLeuGlnAla 87
 Db 148 GTATCTTCAAGTAA-----TCAAG---AACCTTTGTATGAAT 186
 Qy 88 ThrIlePheIleValLeuAlaTyrIleGlnProThr-----GluAla 101
 Db 187 GCTATTCAATATATATATGACAGCTGCACATCACTTGCACAAAGATTAAGAAAGGCA 246
 Qy 102 PheAsp-----LeuGlnIleIleTyrSerAspProGlyTyrHis 115
 Db 247 TTAATTAACATCTCTTAACAAGTCTTGATATATGATCTTAGTTCATGATGACTTTTAC 306
 Qy 116 SerGlnGluAsp----- 119
 Db 307 TCATACAGAAATCAGATGAAAGCACTTGCTTCCGTTGACACCTGAAAGCATTTGCTACT 366
 Qy 120 -----ThrGlnThrLysThrSerSerAsnGlyLysThrPhe 132
 Db 367 CTACACCGTTTATCATCAATTACATTCATCTGAGAAATCAATGCA----- 414
 Qy 133 LeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySerSerAlaGly 152
 Db 415 -----GCMAATTCGAAACCTGAAGTTCCAAAACTGATTTGGTTTCATCTGCAGCA 465
 Qy 153 LeuValSerValValAlaThrSerLeuLeuSerHisPheIle----- 166
 Db 466 ATGACACTGCTGTATGTTGCTGCTTTACT---CATTTATGCTGTTGTTAACTTTCC 522
 Qy 167 -----ProAsnValIleSerThrAsnLysAspIleLeuHis 178
 Db 523 ACTCTTCTGCAGATCAACACCAAGAAATTAAGATTCACAGATTCGATTTGTGAT 582
 Qy 179 AsnValAlaGlnIleAlaHisCysTyrAlaGlnLysLysIleGlySerGlyPheAspVal 198
 Db 583 ATGATAGCTCAAAAGTCCCACTGATTTGCCCAAGTAAAGTGCAGTGCCTTGATGTC 642
 Qy 199 AlaThrAlaIleTyrGlyLeuIleValTyrArgArgPheGlnProAlaLeuIleAsn 217
 Db 643 AGTCTGCTGTATGAGAGTCAGCGCTTATGTTGTTTTCACCAAAAGTCTTCT 699
 RESULT 25
 US-10-128-714-4038
 : Sequence 4038, Application US/10128714
 : Publication No. US20030119013A1
 : GENERAL INFORMATION:
 : APPLICANT: Jiang, Bo
 : APPLICANT: Hu, Wengqi
 : APPLICANT: Tiehkoef, Daniel
 : APPLICANT: Zamudio, Carlos
 : APPLICANT: Eroshkin, Alexey M
 : APPLICANT: Lemieux, Sebastien M
 : TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
 : FILE REFERENCE: 10182-018-999
 : CURRENT APPLICATION NUMBER: US/10/128, 714
 : CURRENT FILING DATE: 2002-04-23
 : PRIOR APPLICATION NUMBER: US 60/285, 697
 : PRIOR FILING DATE: 2001-04-23
 : PRIOR APPLICATION NUMBER: US 60/287, 066
 : PRIOR FILING DATE: 2001-04-27

PRIOR APPLICATION NUMBER: US 60/295,890
 PRIOR FILING DATE: 2001-06-05
 PRIOR APPLICATION NUMBER: US 60/303,899
 PRIOR FILING DATE: 2001-07-09
 PRIOR APPLICATION NUMBER: US 60/316,362
 PRIOR FILING DATE: 2001-08-31
 NUMBER OF SEQ ID NOS: 8603
 SOFTWARE: Patent version 3.1
 SEQ ID NO 4038
 LENGTH: 5158
 TYPE: DNA
 ORGANISM: Aspergillus fumigatus
 US-10-128-714-4038

Alignment Scores:
 Pred. No.: 1,33e-13 Length: 5158
 Score: 213.00 Matches: 61
 Percent Similarity: 38.91% Conservative: 25
 Best Local Similarity: 27.60% Mismatches: 69
 Query Match: 9.55% Indels: 66
 DB: 14 Gaps: 5

US-10-069-062-7 (1-432) x US-10-128-714-4038 (1-5158)

QY 4 AlaPheSerAlaProGlyLysAlaPheLeuAlaGlyLysLeuValLeuGluProIle 23
 DB 4490 GCCTTCCGGCGCGGCAAGGTCCTCCACAGGGGTTATCTGCTCCGACCGCAGC 4549
 QY 24 TyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIle----- 39
 DB 4550 TACACTGGAGCTGTGTCCCTTCAGCCAGATCATGTCATGTCAGCAATTGAGA 4609
 QY 40 -----ThrProLys 42
 DB 4610 CGGAACCATCGCGGAGAGCCGCTCGGATCCGCCATGCGCGGTCGACACACTCGAG 4669
 QY 43 -----GlyThrSerLeuLysGluSerArgIleLysIleSerSerPro 56
 DB 4670 GCGGAGGCAATGTCATGAGAGCAAGAGCAAGAGGACGATCTCGTCACTCCCA 4729
 QY 57 GlnPheAlaSerGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 67
 DB 4730 CAGTTCGTGATCGGATGAGAGTATGATCAATCAAGATCGAGAGCGGTGAGAGATC 4789
 QY 67 ----- 67
 DB 4790 CTAGTGAACAGAGAAATGATGAGTAAGCAACATATTATGATCAATAATGCCCA 4849
 QY 68 SerSerAsnThrGluLysProArgGluValGlnSerArgIleAsnProPheLeuGluAla 87
 DB 4850 GGAAGCGACCATCCAGGCCAGCC-----AACCTGTTTGTGCGAGACC 4891
 QY 88 ThrIlePheIleValLeuAlaTyrIle-----GlnProThrGluAlaPheAsp 103
 DB 4892 TCTCTGAATCCGCTTGACTTACATCAGCTATGCGGACGACATTCGAGGTCGA 4951
 QY 104 LeuGluIleIleIleLysSerAspProGlyTyrHisSerGlnGluAspThrGluThrLys 123
 DB 4952 TTTATGATTAATCTACTCGCGCAACGATTAATCTACTCGAGCGGCTTCCAAAGCT 5011
 QY 124 ThrSerSerAsnGlyLysLysThrPheLeuTyrHisSerArgAlaIleThrGluValGlu 143
 DB 5012 TCGGAGCTCCGCGTCGACAGACATTCGAGACTTTGGTTGCTTCAGAGGAGCAGC 5071
 QY 144 LysThrGlyLeuGlySerSerAlaGlyLeuValSerValValAlaThrSerLeuLeuSer 163
 DB 5072 AAGACAGGCGCTGGCTTCAGCGGCTTGTGATCAGCCCTGTGTCGTCCTGTCATC 5131
 QY 164 His 164
 DB 5132 CAC 5134
 RESULT 26

US-09-815-242-6559
 Sequence 6559, Application us/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 FILE REFERENCE: ELITRA 011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 6559
 LENGTH: 1107
 TYPE: DNA
 ORGANISM: Enterococcus faecalis
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)...(1107)
 US-09-815-242-6559

Alignment Scores:
 Pred. No.: 3.05e-14 Length: 1107
 Score: 209.50 Matches: 113
 Percent Similarity: 39.38% Conservative: 52
 Best Local Similarity: 26.97% Mismatches: 137
 Query Match: 9.39% Indels: 117
 DB: 9 Gaps: 23

US-10-069-062-7 (1-432) x US-09-815-242-6559 (1-1107)

QY 6 SerAlaProGlyLysAlaPheLeuAlaGlyLysLeuValLeuGluProIleTyrAsp 25
 DB 13 ACTAGCCAGAGAAATGTTTATTCAGAGAAATGCGGTGTAACCTGGCCACCT 72
 QY 26 AlaTyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLysLysThrSer 45
 DB 73 GCCATTATCTGTCGCTGAT-----CAATTCGATCTGATCTGAGAGAAACACA 126
 QY 46 LeuLysGluSerArgIleLysIleSerSerProGlnPheAla----- 59
 DB 127 GATGAAGCAGT-----ATTCAATCTGACAAATACAGTCTTACCTATGCTGG 177
 QY 60 -----AsnGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 210
 DB 178 ACACCCGCAAAATGCGAGCTCGTATTAGATATT----- 210
 QY 77 ValGlnSerArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIle 96
 DB 211 -----CGGAAATCTTTT-----CATTAATCTGAGCGGCAAT 246
 QY 97 GlnProThrGluAlaPheAspLeuGluIleIleIleLysSerAspProGlyTyrHisSer 116

```

Db      247 CATCTAACTGAAAAATAT-----GC- 267
Qy      117 GINGLUAEPThrGluThrLysThrSerSerAnglyGluLysPheLeuTyrrHiser 136
Db      268 CAAGAG-----CAAAACAAGAAATTCATTT--TATCATTTA 303
Qy      137 ArgAlaIleThrGluValGlu-----LysThrGlyLeuGlySerSerAla 151
Db      304 AAAGAGACAGATGATAGATTAGTCAAAATGACGAAATATATGCTTGGTCAAGC--- 360
Qy      152 GlyLeuValSerValValAlaThrSerLeuLeuSerHisPheIleProAsnValIleSer 171
Db      361 GGTGACAGTACCGGTGGAACCTGTCAAGCCCTGAATATTTTATGAC-----TTAGGT 414
Qy      172 ThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysGlyThrAlaGlnLysLys 191
Db      415 TTGGAAATAGAGAAATTTTCAAAATATATGACATTAAGCTTACCTTACGCGGAAAT 474
Qy      192 IleGlySerGlyPheAspValAlaThrAlaIleTyrr---GlyLeuIleValTyrrArg 210
Db      475 ---GGTTCTGGGAGATATCGCCGACGCTTACGGGGGCTGATTCCTTTCAACC 531
Qy      211 PheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspProGlyLysPhe 230
Db      532 TTGCATCATGATTGGGTCAAT-----CAAAAGTA 561
Qy      231 ProThrGlu---LeuLysLysLeuIleGluSerAsnTrpGluGlu----- 244
Db      562 ACCACTGAAACATTAACGATTTGTATGACATGACGCTGCAATTAATGATTTTCCG 621
Qy      245 ---LysHISgluArgCysThrLeuProTyrrGlyIleLysLeuLeuMetGlyAspValLys 263
Db      622 TTAATA-----GTACCGAAACAACATACGTTTACTAATGTTGTTGACAGGT 666
Qy      264 GlyGlySerGluThrProLysLeuValSerArgValLeuGlnTrpLysLysGluLysPro 283
Db      667 AGTCCTGGCTGCACCTTCAGACTAGTTCAGTTCATCAATCAAAAGAAAGAAAA--- 723
Qy      284 GluLysSerSerValValTyrrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGlu 303
Db      724 -----CAAGCGGCTTATGACGACGCTTCTTAATGAAA 753
Qy      304 LeuArgGluMetArgGluLysTyrrAspSerAspProGluThrTyrrIleLysGluLeuAsp 323
Db      754 AGTCGGCTTGT-----GTCCGAACAATGATTAAT-----GGC 786
Qy      324 HisSerValGluProLeuThrValAlaIleLysAsnIleArgLysGlyLeuGlnAlaLeu 343
Db      787 TTTAACACAGGAAATTTCTGTATTCAAAAACAATTAATAAATCGCAATTGCTC 846
Qy      344 ThrGlnLysSerGlu-----ValProIleGluProAspValGlnThrGlnLeu 360
Db      847 GCCGATTAATTCCTACGACTGCTGTGTATCGAACAAGAACCTTGAATAATCTTTGT 906
Qy      361 AspArgCysGlnGluIleProGlyCysValGlyGlyValValProGlyValGlyLysTyrr 380
Db      907 GATTTCGCTGAATCTTATACAGAGACTCGGAATCT-----TTCGGCGCTGGCGGGGCGC 960
Qy      381 AspAlaIleValLeuValLeuValLeuGluLysGlnValGlyAsnPheLysGlnLysThr 399
Db      961 GATTTCGCGATTGTATTT-----TTCGGCGCAAAATATCT 993

```

RESULT 27

US-09-070-927A-204

Sequence 204, Application US/09070927A

Patent No. US20020120116A1

GENERAL INFORMATION:

APPLICANT: Charles A. Kunsch

Patrick J. Dillon

Steven Barash

TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 982

```

CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070,927A
FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14

ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 204:
SEQUENCE CHARACTERISTICS:
LENGTH: 6835 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 204:
US-09-070-927A-204

Alignment Scores:
Pred. No.: 5,28e-13 Length: 6835
Score: 209.50 Matches: 113
Percent Similarity: 39.38% Conservative: 52
Best Local Similarity: 26.97% Mismatches: 137
Query Match: 9.39% Indels: 117
DB: 10 Gaps: 23

US-10-069-062-7 (1-432) x US-09-070-927A-204 (1-6835)
Qy      6 SerLAPProGlyLysValPheLeuAlaGlyLysTyrrLeuValLeuGluProIleTyrrAsp 25
Db      3021 ACTACCGCAGGAAAGTATTTATTCAGAGAAATATGCGGTTGTGAACCTGGCCACCT 3080
Qy      26 AlaTyrrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLysGlyThrSer 45
Db      3081 GCCATTATCGTGTGTGGAT-----CAATTGTAACGTAACTGTGGAAGAAACAACA 3134
Qy      46 LeuLysGluSerArgIleLysIleSerSerProGlnPheAla----- 59
Db      3135 GATNAGAGCAGT-----ATTCAATCTGCACAAATACAGCTTTTACATTATCGTTGG 3185
Qy      60 -----AsnGlyLysTrpGluTyrrHisIleSerSerAsnThrGluLysProArgGlu 76
Db      3186 ACAGCGCGAAATGGAGGCTTGTATTAATAT----- 3218
Qy      77 ValGlnSerArgIleAsnProPheLeuGlnAlaThrIlePheIleValLeuAlaTyrrIle 96
Db      3219 -----CGCGAAATTCCTTT-----CATTAATGTTTACCGCGCAT 3254
Qy      97 GlnProThrGluAlaPheAspLeuGluIleIleIleTyrrSerAspProGlyTyrrHiser 116
Db      3255 CATCTAACTGAAAAATAT-----GCG 3275

```

```

QY 117 GINGLUSPSTHRCGLUTHRYSTHRSERASNGLYLULYSTRPHEUTYRHSISER 136
DB 3276 CAAGG-----CAAAACAAAGAAATTCATTT--TATCATTTA 3311
QY 137 ARGALAILETHRCGLUVALGLU-----LYSTHRCGLYLEUGLYSERSERALA 151
DB 3312 AAAGTACAGCAATTAAGTATGCTCAAAATGACGAAATATGCTTTCATTCACAC--- 3368
QY 152 GLYLEUVALSERVALALATHRSERLEUUSERHSPHEILLEPROASNAVALLISER 171
DB 3369 GGTGCGGTACCTTCGAACTGCAACCTTCGAAATTTTATATAC-----TTAGCT 3422
QY 172 THRASNLYASPILEUHSASNAVALAGNILEALHISCYSTYRALAGNULYLS 191
DB 3423 TTGGAATAATGAGAAATTTTCAATATATACAGCATTCCTACCTGACCGTCAAGAAAT 3482
QY 192 ILEGYSERGLYPHEASVALALATHRALALLETYR--GLYLEUVALYTRARGAG 210
DB 3483 ---GGTTCCTTGCGAGATATCGCCGACCTGTTTACGGGGGCTGAGTTGCTTTTCACACC 3539
QY 211 PHEGLNPROALALEUULEASNAVALPHEGLNVALLEUGLUSERASPROGLULYSRPE 230
DB 3540 TTGCATCATGATGGGTCAAT-----CAAAAAGTA 3569
QY 231 PROTHRGU---LEULYSLSLEULLEGUSERASNTGULGU----- 244
DB 3570 GCCACGAAACATTAATCTGATTTGTAGCAATGACCTGACCTGAATTAATGATTTTCCG 3629
QY 245 ---LYSHIGLUARGCYSTHREUPROYRGLYLLEYSLEULEUWECGLYASPVALLYS 263
DB 3630 TTAATA-----GTACCGAAACAACTGCTGTTACTTAATGTTGGACACAGCT 3674
QY 264 GLYGLYSERGLUTHRPROLYSLEUVALSERASVALLEUGLINTRPLYSGLULYSRPE 283
DB 3675 AGTCCTGCGCTCACTCAAGCTTAGTTGATCGAGTCATCATCAAAAGAAAGAAA--- 3731
QY 284 GLUGLUSERSERVALYTRASPGLNLEUASNSERVALASNLEUGLNPHETELYSGLU 303
DB 3732 -----CAAGCGGCTTATGACAGTTCTTAATGANA 3761
QY 304 LEUATRGULMECARGGLULYSTYTRASPSEASPPROGLINTHRYTILLYSGULEUASP 323
DB 3762 AGTCGCGCTTGT-----GTGGAACAAATGATTAAT-----GGC 3794
QY 324 HISSEVALGLUPROLEUTHRVALALAILLYSASNILEARGLYSGILEUGNIALALEU 343
DB 3795 TTTTACACAGAAAAATTTCTGTTATTCAAAAACAAATTAATAAATGCCAATGCTC 3854
QY 344 THRGLNLYSERGLU-----VALPROLLEGLUPROASPVALGLINTHRCGLNLEU 360
DB 3855 GCGAATTAATCTTCACTGACGTGCTGATGAAACAGAACCTTGAAAAATCTTTGCT 3914
QY 361 AAPARGCYSGINGLULLEPROGLCYSVAILGYLVALVALPROGLVALAGLYGLYR 380
DB 3915 GATTGGCTGCAATCTTATACAGAGAGCTGCAATCT-----TCTGCGCTGCGCGGCGC 3968
QY 381 AAPALAILLEALVALLEUVALLEUGLUAENGINVALGLYASNPHELYSGULYSTR 399
DB 3969 GATTGTGGGATTTTAAT-----TTCCGCAAAAATCT 4001

```

```

; PRIOR APPLICATION NUMBER: 60/107,241
; PRIOR FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 21
; LENGTH: 757
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-909-745-21

Alignment Scores:
Pred. No.: 2,966-11
Score: 181.50
Percent Similarity: 45.64%
Best Local Similarity: 31.54%
Query Match: 8.14%
DB: 10 Gaps: 2

US-10-069-062-7 (1-432) x US-09-909-745-21 (1-757)

QY 264 GLYGLYSERGLUTHRPROLYSLEUVALSERASVALLEUGLINTRPLYSGLULYSRPE 283
DB 14 GAGGATCATTCACCTTCATCAATGAGTGTGATCTGAAACAGTGCAGAAAGTCAACCT 73
QY 284 GLUGLUSERSERVALYTRASPGLNLEUASNSERVALASNLEUGLNPHETELYSGLU 303
DB 74 CAGAAATCAAGAGAACATGAGATGAGTAATTTGGGATGCTTAATTCAGTCTTGACAAACCA 133
QY 304 LEUATRGULMECARGGLULYSTYTRASPSEASPPROGLINTHRYTILLYSGULEUASP 323
DB 134 CTGAGAGAACCTTAACAACTGCTGAMAGATCATGGAAGAGCTATGAATCTGTTTA--- 190
QY 324 HISSEVALGLUPROLEUTHRVAL-----ALAILLYSASNILEARGLYSGLY 339
DB 191 CGATCTGATGCTGCTCACTGACGTAAGTACAGACAGAGTGTGCTTCAATCAATCA 250
QY 332 -----ALAILLYSASNILEARGLYSGLY 339
DB 251 GAATTAATGTTAGATCATTACTGCGCCAGAGATGCTTCTTGAATTAAGCTTCAT 310
QY 340 LEUGNIALALEUTHRCGLNLYSERGLUVALPROLLEGLUPROASPVALGLINTHRCGLN 359
DB 311 ATGCAAGAGATGCGCATGACGCTGCTGTTCAATTAAGCCAGATCACAAACTCACTT 370
QY 360 LEUASPARGCYSGINGLULLEPROGLCYSVAILGYLVALVALPROGLVALAGLYGLY 379
DB 371 CTGAGTGCACCTATGAATATGAGGCTGTTCTACTAGCTGAGATTCTGCGCGCGTGC 430
QY 380 TYRAAPALAILLEALVALLEUVALLEU 388
DB 431 TTTGATGCAAGTTTCTCATGATTTTG 457

RESULT 29
US-09-815-242-4799
; Sequence 4799, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl H.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848

```

;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 4799
;; LENGTH: 1035
;; TYPE: DNA
;; ORGANISM: Staphylococcus aureus
US-09-815-242-4799

Alignment Scores:
Pred. No.: 4,68e-10 Length: 1035
Score: 173.00 Matches: 97
Percent Similarity: 36.24% Conservative: 57
Best Local Similarity: 22.82% Mismatches: 143
Query Match: 7.76% Indels: 128
Gaps: 20
DB: 9

US-10-069-062-7 (1-432) x US-09-815-242-4799 (1-1035)

QY 7 AlaProGlyLysAlaPheLeuAlaGlyLysLeuValLeuGluProIleTyrAspAla 26
DB 7 GCACCCGGAACCTTATATCTGCGAATATGCTGAACAGAACAGCAATATATAACT 66
QY 27 TyrValThrAlaLeuSerSerArgMetHisAlaValIleThrPro-----Lys 42
DB 67 GTACTTATTCGCTTGATGCTTTTGTGTAAGCTGCTACTATTGAGAAACAGCAATATATA 126
QY 43 GlyThr-----SerLeuLysGluSer 49
DB 127 GGTACCATTCATCAAAAGCATTAATCAATCAACCATTAATTAATTAATTAATTAATTA 186
QY 50 ArgIleLysLeuSerSerProGlnPheAlaAsnGlyLysTyrGluTyrHisLeuSer 69
DB 187 AGTATTGTCATTCAGATTCATGACATGACGCAAAACA----- 222
QY 70 AsnThrGluLysProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThrIle 89
DB 223 -----TTAATTATGTCGTC 237
QY 90 PheIleValLeuAlaTyrIleGlnProThrGluAlaPheAspLeuGluIleIleTyr 109
DB 238 ACGCTATTGAATTTTGAACATACGGAAGTTGGATATAGCCATG----- 288
QY 110 SerAspProGlyTyrHisSerGlnLysPheThrGluTyrLysSerSerAsnGlyLys 129
DB 289 -----AAGCATTTTCATGCTGACTATGTAGTAATTTA-----GATGATTCAATGCT 336
QY 130 LysThrPheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySer 149
DB 337 -----CATTAATATGATGATTGATCA 357
QY 150 SerAlaGly-----LeuValSerValAlaAlaThrSerLeuLeuSerHisPheLeuProAsn 168
DB 358 AGTCAGACGACTCTGTGTCAGTTATA-----AAAGTATTAAGAAATTTATGATATG 411
QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
DB 412 AAGTATCTAATTTA-----TACATTTATAAAGTAGAGATTTGCAAAATGAAATTA 465
QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyr-----GlyLeuIleVal 207
DB 466 CAAGATTAAAGTTATGCGGA-----GATATGCTGTGAGATGATATATAGATGCTACG 522

QY 208 TyrArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspPro 227
DB 523 TATAGTACTTTGATCATGAATGGGTAAAGCATCAATTAATGA----- 564
QY 228 GlnLysPheProThrGlnLeuLysLysLeuIleGluSerAsnProGluLysHis--- 246
DB 565 -----GATCTACGCTTGAGAAAGATTTTAATCAAAAACCTGCGATTTGACATC 615
QY 247 GlnArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysGlyLysSer 266
DB 616 GAACCATTAACAAGCACTGAAATATGGAAGTACTTATCGGTGACGCTGACCCGCGC 675
QY 267 GlnThrProLysLeuValSerArgValLeuGlnTyrLysGlyLysProGluLysSer 286
DB 676 TCATCACCACACTTTGTTAGGGAAGTGAACGTTTGAA----- 714
QY 287 SerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGluLeuArgLys 306
DB 714 ----- 714
QY 307 MetArgGluLysTyrAspSerAspProGluThrTyr-----IleLysGluLeuAsp 323
DB 715 -----TGGATCTTCATTTTACGGTGCCTTCTTGAAGATTCACAT 756
QY 324 HisSerValGluProLeuThrValAlaIleLys---AsnIleArgLysGlyLeuGlnAla 342
DB 757 CGTGTGTTGAAAACCTTATTCATGCTTTTAAACAAATTAACATTAAGGTGCAAAAG 816
QY 343 LeuThrGlnLysSerGluValProIleGlu-----ProAspVal 355
DB 817 ATGGCGCTGCAAGATCGTACATTAATTCAGATGATGATAAAGACATCAAGTTGATATA 876
QY 356 GlnThrGln-----LeuLeuAspArgCysGlnGluIleProGlyCysValGly 371
DB 877 GAACCTGAAAAGCTTAATATTTGTGTATTTGCTGAAAAGTATCACGCG-----GCA 930
QY 372 GlyValValProGlyAlaGlyLysTyrAspAlaIleAlaValLeuValLeuGluAsnGln 391
DB 931 TCTAAACATCAGAGCGCGTGGTGGAGACGCT---GGTATTACATTAATCAATTAAGAT 987
QY 392 ValGlyAsnPheLys 396
DB 988 GTAGATAAAGAAAA 1002

RESULT 30
US-09-815-242-8768
; Sequence 8768, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FASTSEQ for Windows Version 4.0
 SEQ ID NO 8768
 LENGTH: 1077
 TYPE: DNA
 ORGANISM: Staphylococcus aureus
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)...(1077)
 US-09-815-242-8768

Alignment Scores:
 Pred. No.: 4,98E-10 Length: 1077
 Score: 173.00 Matches: 97
 Percent Similarity: 36.24% Conservative: 57
 Best Local Similarity: 22.82% Mismatches: 143
 Query Match: 7.76% Indels: 128
 DB: 9 Gaps: 20

US-10-069-062-7 (1-432) x US-09-815-242-8768 (1-1077)

QY 7 AlaProGlyLysAlaPheLeuAlaGlyGlyTyrLeuValLeuGluProIleTyrAspAla 26
 DB 16 GAACCCGGAACCTTATATGCTGAGAAATATGCTAACAGAACGAGATATTAATCT 75
 QY 27 TyrValThAlaLeuSerSerArgMetHisAlaValIleThrPro-----Lys 42
 DB 76 GATCTATTGCGTTAGATCGTTTGTACGCTACTACTTGAAGAAACAGCAATATTA 135
 QY 43 GlyThr-----SerLeuysGluSer 49
 DB 136 GGTACCATTCATTCAAAGCATTAATCATACCCGCTTACATTTGTAGATGAAGAT 195
 QY 50 ArgIleLysIleSerSerProGlnPheAlaAsnGlyLutPgluTyrHisSerSer 69
 DB 196 AGATTGTGATTTCAATCCATGATCCAGCAAGCAAAACA----- 231
 QY 70 AsnThrGluLysProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThrIle 89
 DB 232 -----TTAAATTAATGCGTC 246
 QY 90 PheIleValLeuAlaTyrIleGlnProThrGluAlaPheAspLeuIleIleTyr 109
 DB 247 AAGCATTAATGAATATTTGAACATATCGGAAAGTTGCGATATACCATG----- 297
 QY 110 SerAspProGlyTyrHisSerGlnGluAspThrGluThrLysThrSerSerAsnGlyGlu 129
 DB 298 -----AAGCATTTTCATCGATATGATATTAATTTA-----CATGATTCAAATGCT 345
 QY 130 LysThrPheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySer 149
 DB 346 -----CATTAATTAATGATTAAGCTTCA 366
 QY 150 SerAlaGly-----LeuValSerValValAlaThrSerLeuLeuSerHisPheProan 168
 DB 367 AGTGGCAGCATTTGTGTGAGTTATA-----AAAGTATTAATTAATTTATGATGATG 420
 QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
 DB 421 AAGTATCTAATTTA-----TACATTTTAATTAACGAGTATTCGAATATGAAGTTA 474
 QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyr--GlyLeuIleVal 207
 DB 475 CAAAGTTTAAGTTCAATGCGGA-----GATATTGCTGTAGATATATAGTGAATGCTAAG 531
 QY 208 TyrArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspPro 227
 DB 532 TATAGTACTTTTGCATGATGAATGAGTTAAGCATCAATATGA----- 573
 QY 228 GluLysPheProThrGluLeuLysLysLeuIleGluSerAsnTrpGluLysHis--- 246

DB 574 -----GATACACGCTTGAAAGAGCTTTTAATCAAAAACCTGGCTGATTCACATC 624
 QY 247 GlnArgCysThrLeuProTyrGlyIleLysLeuMetGlyAspValLysGlySer 266
 DB 625 GAACCATTAACAAGCACTGAAATATGGAAGTACTTATCGTTGACGCTCACCGGCG 684
 QY 267 GlnThrProLysLeuValSerArgValLeuGlnTrpLysGlyLysProGluLysSer 286
 DB 685 TCATCACCACTTGTATTAGGCAAGTGAACGTTGAAA----- 723
 QY 287 SerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGluLeuArgGlu 306
 DB 723 ----- 723
 QY 307 MetArgGluLysTyrAspSerAspProGluThrTyr-----IleLysGluLeuAsp 323
 DB 724 -----TCATATCTTCATTTACGGTGACTTCTTGAAATTCACAT 765
 QY 324 HisSerValGluProLeuThrValAlaIleLys--AsnIleArgLysGlyLeuGlnAla 342
 DB 766 CGTTGTGTGAAGAACTTATTCATGCTTTTAAACAAATTAACATTAAGTGTGCAAAAG 825
 QY 343 LeuThrGlnLysSerGluValProIleGlu-----ProAspVal 355
 DB 826 ATGTGCGCTCAAGATCGTACATTAATTCACGATGATGAAGAACTACAGTTGATATA 885
 QY 356 GlnThrGln-----LeuLeuAspArgCysGlnGluIleProGlyCysValGly 371
 DB 886 GAACCTGAAAGCTAAATATTGTGTATTTGCTGTAAGATATACAGG-----GCA 939
 QY 372 GlyValValProGlyValGlyTyrAspAlaIleAlaValLeuValLeuGluAsnGln 391
 DB 940 TCTAAACATCAACGGCGCTGCTGTGAGACATGCT---GGTATTACATTAATTAAGAT 996
 QY 392 ValGlyAsnPheLys 396
 DB 997 GTAGATTAAGAAAAA 1011

RESULT 31

US-08-781-986A-243/C
 Sequence 243, Application US/08781986A
 Publication No. US20030054436A1
 GENERAL INFORMATION:
 APPLICANT: Charles Kunsch
 TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
 NUMBER OF SEQUENCES: 5255
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/781,986A
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Benson, Bob
 REGISTRATION NUMBER: 30,446
 REFERENCE/DOCKET NUMBER: PB248PP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 243:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 10146 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 US-08-781-986A-243

Alignment Scores:
 Seq. No.: 1,676-08 Length: 10146
 Score: 173.00 Matches: 97
 Percent Similarity: 36.24% Conservative: 57
 Best Local Similarity: 22.82% Mismatches: 143
 Query Match: 7.76% Indels: 128
 Gaps: 20

US-10-069-062-7 (1-432) x US-08-781-986A-243 (1-10146)

QY 7 AlaProGlyLysAlaPheLeuAlaGlyGlyTyrLeuValLeuGluProIleTyrAspAla 26
 DB 3516 GCACCCGGAACCTTATATCTGCGAATATGCTGTAACAGAACCGATATTAATCT 3457
 QY 27 TyrValThrAlaLeuSerSerArgMetHisAlaValIleThrPro-----Lys 42
 DB 3456 GTACTTATTCGCTGATCGTTTGTGACTGCTACTATTGAAGACACCAATATATA 3397
 QY 43 GlyThr-----SerLeuylsGluSer 49
 DB 3386 GGTACCATTCATTCAAAAGCATTAACATCCAGTTACATTAGTAGATTAAGAT 3337
 QY 50 ArgIleLysIleSerSerProGlnPheAlaenGlyGluTyrGluTyrHisIleSerSer 69
 DB 3336 AGTATTGTCATTTGATGATCAGATCAGACGAAACAA----- 3301
 QY 70 AsnThrGluLysProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThrIle 89
 DB 3300 -----TTAATTATGTCGTC 3286
 QY 90 PheIleValLeuAlaTyrIleGlnProThrGluAlaPheAspLeuGluIleIleTyr 109
 DB 3285 ACAGCATTTGAATTTTGAACAAATACCGGAAAGTGGATATACCGATG----- 3235
 QY 110 SerAspProGlyTyrHisSerGlnGlnAspThrGluThrLysThrSerSerAsnGlyGlu 129
 DB 3234 -----AAGCATTTCACTGATGATGATGATTA-----GATGATTTCAAAATGT--- 3187
 QY 130 LysThrPheLeuTyrHisSerArgAlaIleThrGluValGluLysThrGlyLeuGlySer 149
 DB 3186 -----CAATAATGATGATTAAGTTCA 3166
 QY 150 SerAlaGly-----LeuValSerValAlaIleThrSerLeuLeuSerHisPheIleProAsn 168
 DB 3165 AGTGCAGCAATGATCTGTGTCAGTTATA-----AAAGTATTAAAGAAATTTATGATATG 3112
 QY 169 ValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAla 188
 DB 3111 AAGTATCTAATTA-----TACATTATAAAGTACAGATGATGCAAAATATGAAGTGA 3058
 QY 189 GlnLysLysIleGlySerGlyPheAspValAlaThrAlaIleTyr-----GlyLeuIleVal 207
 DB 3057 CAAGTTTAAGTTTACATGCGGA-----GATATTCCTGATGATATATATGATGATGCTACCG 3001
 QY 208 TyrArgAspPheGlnProAlaLeuIleAsnAspValPheGlnValLeuGluSerAspPro 227
 DB 3000 TATAGTACTTTGATCATGATGATGGGTAAAGCATCAAAATTGAA----- 2959
 QY 228 GluLysPheProThrGluLeuLysLysLysLeuIleGluSerAsnTyrGluLysHis--- 246
 DB 2958 -----GATCTACAGCGTTGAAGAGATTTTAATCAAAAACCTGCGCGATGACATC 2908
 QY 247 GluLysGlyThrLeuProTyrGlyLysLeuLeuMetCysIleAspValLysGlySer 266
 DB 2907 GAACCATTAACAAGCACTGAAATATGAAAGTACTTATCGGTGATGCTGCACCGCGC 2848

QY 267 GluThrProLysLeuValSerArgValLeuGlnIleTyrLysGluLysProGluGluSer 286
 DB 2847 TCATCACCACTTTGTTATGCAAGTGAAGCTTTGAAA----- 2809
 QY 287 SerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGluLeuArgGlu 306
 DB 2809 ----- 2809
 QY 307 MetArgGluLysTyrAspSerAspProGluThrTyr-----IleLysGluLeuAsp 323
 DB 2808 -----TCAGATTCCTTCATTTTAACGTCGATCTTTAGAAAGATTACAT 2767
 QY 324 HisSerValGluProLeuThrValAlaIleLys-----AsnIleArgLysGlyLeuGlnAla 342
 DB 2766 CGTTGCTGTGAAAACCTTATTCATGCTTTTAAACAAATTAATTAAGGTGTGAAAG 2707
 QY 343 LeuThrGlnLysSerGluValProIleGlu-----ProAspVal 355
 DB 2706 ATGGTGCCTGCAAGATCGTACAAATTAATTCAGATGATTAAGAAAGCTACAGTTGATATA 2647
 QY 356 GlnThrGln-----LeuLeuAspArgCysGlnGluIleProGlyCysValGly 371
 DB 2646 GAACCTGAAAAGCTAAATATTTGTGTGATTTGCTGAAAAGTATCACGGC-----GCA 2593
 QY 372 GlyValValProGlyAlaGlyLysTyrAspAlaIleAlaValLeuValLeuGluAsnGln 391
 DB 2592 TCTAAACATCAAGCGCGCTGTGTGTGAGACTGT---GGTATTACAAATTATCAATTAAGAT 2536
 QY 392 ValGlyAsnPheLys 396
 DB 2535 GTAGATTAAGAAAAA 2521

RESULT 32
 US-09-918-740-57
 ; Sequence 57, Application US/09918740
 ; Publication No. US20030033626A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hahn, Frederick
 ; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
 ; FILE REFERENCE: KAS-103XC1
 ; CURRENT APPLICATION NUMBER: US/09/918, 740
 ; PRIOR FILING DATE: 2001-07-31
 ; PRIOR APPLICATION NUMBER: 60/221,703
 ; NUMBER OF SEQ ID NOS: 76
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 57
 ; LENGTH: 6798
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Streptomyces sp. CL190 gene cluster containing mevalonate pathway
 US-09-918-740-57

Alignment Scores:
 Seq. No.: 9,246-06 Length: 6798
 Score: 147.00 Matches: 100
 Percent Similarity: 36.10% Conservative: 61
 Best Local Similarity: 22.42% Mismatches: 161
 Query Match: 6.59% Indels: 124
 Gaps: 23

US-10-069-062-7 (1-432) x US-09-918-740-57 (1-6798)

QY 7 AlaProGlyLysAlaPheLeuAlaGlyGlyTyrLeuValLeuGluProIleTyrAspAla 26
 DB 2241 GCAGCCGGAAGCTGTTGTCGCGGAGAGTACCGGATGCGGAGTCCGCGCAACCCGCGC 2300
 QY 27 TyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLysGlyThrSerLeu 46

```

Db      2301 ATCTGTGATGACCGGACATCATGTCACCGTCCGACCGGACGCGACACC 2360
      :|||:
      :|||:
      :|||:
Qy      47 LysGluSerArgIleLeuSer-----ProGlnPheAlaSerGluTyrP 63
      :|||:
      :|||:
      :|||:
Db      2361 GGGGCGCGGACGTCGTGATCTCTCCGACCTCGGCGGACGCGTCCGAC-----TGG 2414
      :|||:
      :|||:
      :|||:
Qy      64 GluTyrHis-----IleSerSerThrGluPheProArgGluValGlnSer 79
      :|||:
      :|||:
      :|||:
Db      2415 CGCTGGCAGACGCGCGGCTCTCTCGGACCGGACGCGGACGCGGACGCGGACG 2474
      :|||:
      :|||:
      :|||:
Qy      80 ArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnProThr 99
      :|||:
      :|||:
      :|||:
Db      2475 GGCCTGGCCACGCGTGTGCGGATC----- 2501
      :|||:
      :|||:
      :|||:
Qy      100 GluAlaPheAsnLeuGluIleIleIleTyrSerAspProGluTyrHisSerGlnGluAsp 119
      :|||:
      :|||:
      :|||:
Db      2502 -----GAGACC 2507
      :|||:
      :|||:
      :|||:
Qy      120 ThrGluThrIleThrSerSerAsnGluGluTyr-----ThrPheLeuTyrHis 135
      :|||:
      :|||:
      :|||:
Db      2508 GTGGCGGCGTGTGGGCGGACCGGACGAGAGTCCCGCTTCACCTCTCCGTCAC 2567
      :|||:
      :|||:
      :|||:
Qy      136 SerArgAlaIleThrGluValGluTyrThrGlyLeuGlySerSerAlaGlyLeuValSer 155
      :|||:
      :|||:
      :|||:
Db      2568 ACCCGGCTGACAGAGGACGCGCGGAGTTCGGCTGGCTCCGACG-----GGCGGCGTAC 2624
      :|||:
      :|||:
      :|||:
Qy      156 ValValAlaThrSerLeuLeuSerHisPheIleProAsnValIleSerThrAsnLysAsp 175
      :|||:
      :|||:
      :|||:
Db      2625 GTGGGACCGTACCGCGCGTGTGCGGATCTGGGACGAGTTCAC-----GAC 2678
      :|||:
      :|||:
      :|||:
Qy      176 IleLeuHisAsnValAlaGlnIleAlaHisCysTyrAlaGlnLysIleGlySerGly 195
      :|||:
      :|||:
      :|||:
Db      2679 GAACGGTTCGCGTGGCCATGCTCCGACCGCGGACCTGACCCCAAG-----GGCTCCGCG 2735
      :|||:
      :|||:
      :|||:
Qy      196 PheAspValAlaThrAlaIleTyr--GlyLeuIleValTyrArgArgPheGlnProAla 214
      :|||:
      :|||:
      :|||:
Db      2736 GGGGACCTCGCGCGGACACCTGGGCGGCTGATGCTACACGCGCGCGCGCGCGC 2795
      :|||:
      :|||:
      :|||:
Qy      215 LeuIleAsnAspValPheGlnValLeuGluSerAspProGluLysPheProThrGluLeu 234
      :|||:
      :|||:
      :|||:
Db      2796 TTGTGCTGACCTGGCGCGCGCGGAGTCCGAC----- 2831
      :|||:
      :|||:
      :|||:
Qy      235 LysLeuLeuIleGluSerAsnTyrGlu--GluLysHisGluArgCysThrLeuProTyr 253
      :|||:
      :|||:
      :|||:
Db      2832 ----CGGACACTGAAGCGCGCTGGCGGCGGACCTCGGCGGACCTCGGCGGCGGCGG 2888
      :|||:
      :|||:
      :|||:
Qy      254 GlyIleLeuLeuLeuMetGlyAspValLysGlyIleSerGluThrProLysLeuValSer 273
      :|||:
      :|||:
      :|||:
Db      2889 GGCCTCACCTGAGGTGGCTGGACCGGAGACCGCGCTCCACCGCTCCGCTGCTGCTCC 2948
      :|||:
      :|||:
      :|||:
Qy      274 ArgValLeuGln-----TyrLysLysGluLysProGluGluSerVal-----Val 289
      :|||:
      :|||:
      :|||:
Db      2949 GATCTCACCGCGGACCTGGCGGCGGACGCGCTCCACCAAGGTTCTGTCGAGACCACG 3008
      :|||:
      :|||:
      :|||:
Qy      290 TyrAspGlnLeuAsnSerAla-----AsnLeuGlnPheMet 301
      :|||:
      :|||:
      :|||:
Db      3009 ACCGACTGTGTGCGCTCGGCGTCCAGCGCGCTCGAGTCCGCGGACGACGACGCGCTG 3068
      :|||:
      :|||:
      :|||:
Qy      302 LysGluLeuArgGluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGlu 321
      :|||:
      :|||:
      :|||:
Db      3069 CACGAGATCCGCGCGCGCGCGGACGAG-----CTGGCGCGCG 3104
      :|||:
      :|||:
      :|||:
Qy      322 LeuAspHisSerValGluProLeuThrValAlaIleLysAsnIleArgLysGlyLeuGln 341
      :|||:
      :|||:
      :|||:
Db      3105 CTGGACGACGAGGTC-----GGCTTCGCGC 3128
      :|||:
      :|||:
      :|||:
Qy      342 AlaLeuThrGlnLysSerGluValProIleGluProAspValGlnThrGlnLeuAsp 361
      :|||:
      :|||:
      :|||:
Db      3129 ATCTTCACACCAAG-----CTGACGCGCGCTGTCGAC 3161
      :|||:
      :|||:
      :|||:
Qy      362 ArgCysGlnGluIleProGluCysValGlyGlyValVal-----ProGluAlaGlyGly 379
      :|||:
      :|||:
      :|||:

```

```

Db      3162 GCGCGCGAA-----GCCGTGCGCGCGCGCGGACCCCTCCGCGGACGCGGC 3209
      :|||:
      :|||:
      :|||:
Qy      380 TyrAsp-----AlaIleAlaValLeuValLeuGlu-----AsnGlnValGlyAsnPheLys 396
      :|||:
      :|||:
      :|||:
Db      3210 GCGCGTGGCGGACGCTGCTGCGGACCGCGGCGGCTGCGGACATCACATGTACG 3269
      :|||:
      :|||:
      :|||:
Qy      397 GlnLysThrLeuGluAsnProAspTyrPheHisAsnValTyrTyrValAspLeuGluGlu 416
      :|||:
      :|||:
      :|||:
Db      3270 CAACGG-----TGG----- 3278
      :|||:
      :|||:
      :|||:
Qy      417 GlnThrGluGlyValLeu 422
      :|||:
      :|||:
      :|||:
Db      3279 GAGACAGCGCGGCTGCTG 3296
      :|||:
      :|||:
      :|||:
RESULT 33
US-09-918-740-63
/ Sequence 63, Application US/09918740
/ Publication No. US2003003626A1
/ GENERAL INFORMATION:
/ APPLICANT: Hahn, Frederick
/ TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
/ TITLE OF INVENTION: create novel traits in transgenic organisms
/ FILE REFERENCE: KAS-103XCI
/ CURRENT FILING DATE: 2001-07-31
/ PRIOR APPLICATION NUMBER: 60/221,703
/ PRIOR FILING DATE: 2000-07-31
/ NUMBER OF SEQ ID NOS: 76
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 63
/ LENGTH: 8077
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Operon F containing A. thaliana, S. cerevisiae, and Streptomyces
/ OTHER INFORMATION: C1190
/ OTHER INFORMATION: DNA
US-09-918-740-63

Alignment Scores:
Pred. No.: 1,21e-05 Length: 8077
Score: 147.00 Matches: 100
Percent Similarity: 36.10% Conservative: 61
Best Local Similarity: 22.42% Mismatches: 161
Query Match: 6,59% Indels: 124
DB: 11 Gaps: 23

US-10-069-062-7 (1-432) x US-09-918-740-63 (1-8077)
Qy      7 AlaProGlyLysValAlaPheLeuAlaGlyLysTyrLeuValLeuGluProIleTyrAspAla 26
      :|||:
      :|||:
      :|||:
Db      3484 GCGCGGCGGACCTGCTGCGGCGGACGAGTACGCGCTGCTGATCCGCGGACACCGGCG 3543
      :|||:
      :|||:
      :|||:
Qy      27 TyrValThrAlaLeuSerSerArgMetHisAlaValIleThrProLysGlyThrSerLeu 46
      :|||:
      :|||:
      :|||:
Db      3544 ATCTGTGATGACCGGACATCATGTCACCGTCCGACCGGACGCGGACGACACC 3603
      :|||:
      :|||:
      :|||:
Qy      47 LysGluSerArgIleLeuSer-----ProGlnPheAlaSerGluTyrP 63
      :|||:
      :|||:
      :|||:
Db      3604 GGGGCGCGGACGTCGTGATCTCTCCGACCTCGGCGGACGCGTCCGCGC-----TGG 3657
      :|||:
      :|||:
      :|||:
Qy      64 GluTyrHis-----IleSerSerThrGluLysPheProArgGluValGlnSer 79
      :|||:
      :|||:
      :|||:
Db      3658 CGCTGGCAGACGCGCGGCTGCTCGGCGGACCGGACGAGGACGAGGCGGCGGACG 3717
      :|||:
      :|||:
      :|||:
Qy      80 ArgIleAsnProPheLeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnProThr 99
      :|||:
      :|||:
      :|||:
Db      3718 GGCCTGGCCACGTCGTGCGGATC----- 3744
      :|||:
      :|||:
      :|||:
Qy      100 GluAlaPheAsnLeuGluIleIleIleTyrSerAspProGluTyrHisSerGlnGluAsp 119
      :|||:
      :|||:
      :|||:
Db      3745 -----GAGACC 3750
      :|||:
      :|||:
      :|||:

```


Oy	120	ThrluThrluThyThrsSerSerhngllyuvs-----ThPheleuThyHls	135
Db	3751	GTGGGCGGGCTCTGGGCGAAGCGGACGAGAAAGTCCCGGCTCTACCTCTCCGCTACG	3811
Oy	136	SerATgAlAlleThrGluValGluTyThrGlyLeuGlySerSerAlaGlyLeuValSer	155
Db	3811	AGCCGCTGTACAGAGAGACGGCCGGAAGTTCCGCTCTGGCTTCAGC---GGCGGGTGAAC	3861
Oy	156	ValValAlAlThrSerLeuLeuSerhIbPheIleProAsnValIleSerThAsnlyAsp	175
Db	3868	GTGGGACCGTGAAGCCGCGCTGCAGCGCTTCTGGCATCGAACTGTCAAC-----GAC	3921
Oy	176	IleLeuHISAsnValAlaGlnIleAlaHisCysTyrAlaGlnTyLeuIleGlySerGly	195
Db	3922	GAACTGTCGGGCTGGCTGGCATGTCCGCTCCACCGCGAACTGACCCCAAG---GGCTTCGGG	3976
Oy	196	PheAspValAlAlAlThraAlAlleTyr---GlyLeuIleValTyrArgArgPheGlnProAla	214
Db	3979	GGGAGACTTCGCGCGGACGACCTGGGGGCGGCTGATCGCTTACAGCGCGCGGACCGGGCG	4033
Oy	215	LeuIleAsnAspValPheGlnValLeuGlnSerAspProGlnTyLePheProThrGluLeu	234
Db	4039	TTTGTGCTCGACCTGGCCCGGCGCGCTGGAGTGCAC-----	4074
Oy	235	LySlySLeuIleGlySerSerAsnTrpGln---GlnTySNIsgLuarGysThrLeuProTyr	253
Db	4075	---CGGACACTGAAGCGCCGCTGGCGGGGACCTGGTGGCGGCACTCGCGGCGCCAAAG	4133
Oy	254	GlyTyleuLeuMetGlyAspValTySglyTySergTyrThrProLyLeuValSer	273
Db	4132	GGCTCTACCCCTGGAGTGGCTGGACCGGAGACCGCGCTCCACCGGCTCTGGTGTCTC	4191
Oy	274	ArgValLeuGln-----TrpLySlySglnTySProGlnGlySerSerVal-----Val	289
Db	4192	GATCTGACCGCGCGACCTGGCGGGGACAGCGCTCCACCAAGAGTTGTCAGACCAAG	4251
Oy	290	TyrAspGlnLeuAsnSerAla-----AsnLeuGlnPheMet	301
Db	4252	ACCGACTGTGTCCGCTCGCGGTCACGCGCTCGAGTCCGGGACAGACAGACGCTGCTG	4311
Oy	302	LySglnLeuAspGlnMetArgGlnTySArgSerAspProGlnThrTyrIleTySgln	321
Db	4312	CACGAGATCCGCGCGGCGCCGCGCAGAG-----CTGGCCCGCG	4347
Oy	322	LeuAspHISerValGluProLeuThrValAlAlleTySAsnIleArgLyGlnGln	341
Db	4348	CTGGACGACGAGTCT-----GGCTTCGGC	4371
Oy	342	AlaLeuThrGlnTySserGlnValProIleGluProAspValGlnThrGlnLeuAsp	361
Db	4372	ATCTTCACACCCAAAG-----CTGACGGCGCTGTGGCAG	4404
Oy	362	ArgCysGlnGlnIleProGlyCysValGlyTyValVal-----ProGlyAlaGlyTy	379
Db	4405	GCCCGCGAA-----GCCCTCGCGCGCGCGGCAAGCCTCTCGGGGCAAGCGCGC	4452
Oy	380	TyrAsp---AlAlleAlaValLeuValLeuGln-----AsnGlnValGlyAsnPheTyS	396
Db	4453	GGCAGCTGGGCACTGCCCTGTGTGACCGCCAGAGGCTGGCGGACATACACATGTACGG	4512
Oy	397	GlnTySThrLeuGlnAsnProAspTyrPheHISAsnValTyrTrpValAspLeuGlnGln	416
Db	4513	CAACGG-----TGG-----	4521
Oy	417	GlnThrGlnGlyValLeu	422
Db	4522	GAGAGACCGCGGAGTGTCTG	4539

RESULT 34
US-09-909-745-19
; Sequence 19, Application US/09909745
; Patent No. US20020119546A1

```

/ GENERAL INFORMATION:
/ APPLICANT: Falco, S. Carl
/ TITLE OF INVENTION: Squalene Synthesis Enzymes
/ FILE REFERENCE: Bb112 US CIP
/ CURRENT APPLICATION NUMBER: US/09/909,745
/ PRIOR FILING DATE: 2001-07-20
/ PRIOR APPLICATION NUMBER: 60/107,241
/ PRIOR FILING DATE: 1998-11-05
/ PRIOR APPLICATION NUMBER: 60/107,241
/ PRIOR FILING DATE: 1998-11-05
/ NUMBER OF SEQ ID NOS: 26
/ SOFTWARE: Microsoft Office 97
/ SBO ID NO 19
/ LENGTH: 539
/ TYPE: DNA
/ ORGANISM: Oryza sativa
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (287)
/ NAME/KEY: unsure
/ LOCATION: (440)
/ NAME/KEY: unsure
/ LOCATION: (454)
/ NAME/KEY: unsure
/ LOCATION: (460)
/ NAME/KEY: unsure
/ LOCATION: (465)
/ NAME/KEY: unsure
/ LOCATION: (486)
/ NAME/KEY: unsure
/ LOCATION: (488)
/ NAME/KEY: unsure
/ LOCATION: (495)
/ NAME/KEY: unsure
/ LOCATION: (530)
/ NAME/KEY: unsure
/ LOCATION: (533)
/ US-09-909-745-19

Alignment Scores:
Pred. No.:          9 87e-07          Length:          539
Score:             140.50          Matches:          46
Percent Similarity: 43.26%          Conservative:     15
Best Local Similarity: 32.62%          Mismatches:      54
Query Match:        6.30%           Indels:          27
DB:                 10              Gaps:            2

US-10-069-062-7 (1-432) x US-09-909-745-19 (1-539)
Qy      264  GLVGLYsercgluThrProLysleuValSerArgValleuGlnTrpLysGluLysPro 283
Db      7    GGAGGATCATCCACTCCATCATGTTGGATCTGTGGAACAGTGGCAGAACTCAGACCT 66
Qy      284  GluGluSerSerValValIyrAspGlnLeuAsnSerAlaasnleuGlnPheMetLysGlu 303
Db      67    CAGAAATCCAGAGACATGAGGTAAATTGGGATTTGCTAATTCAGTGCCTTGAGAACAA 126
Qy      304  LeuArgGluMetArgGluLysIyrAspSerAspProGluThrIyrTrileYsGluLeuAsp 323
Db      127  CTAGGAACTTAACAACTTGCTGAGATCATCTGGGAAAGCCTATGAAATCTGTTTA--- 183
Qy      324  HisSerValGluProleuThrVal-----AlaIleYsasnIleArgLysGly 339
Db      184  CGATCTCTGTAAGTCGTCTCACTGTCAGTAAGTGAAGAGAGTGGCTACCAATCAACATCAA 243
Qy      332  -----
Db      244  GAACATAATTGTAGATCATTAATGAGCCGCAAGAGATGCTTCCNTGAAATPAGGCTTCAT 303
Qy      340  Leu-GlnAlaLeuThrGlnLysSerGluValProIleGluProAspValGlnThrGluLe 359
Db      304  ATGCCAAGAGATGGGCAATGACAGCTGGTTCATTAATGAGCCAGAAATTCACAACTCAACT 363

```

QY 359 ULEUAPARGYGLNGLUILEPROGLYCYSEVALGLYVALVALPROGLYVALAGLY 379
DB 364 TGTGATGCCAC-TATGAAATATGAGAGGTCTTCTACTACTGATTCCTGGGCGCGTGG 422

QY 379 Y 379
DB 423 C 423

RESULT 35

US-09-988-863A-5
Sequence 5, Application US/0998863A
Patent No. US20020123427A1
GENERAL INFORMATION:
APPLICANT: Bayer AG
TITLE OF INVENTION: Plant phosphomevalonate kinases
FILE REFERENCE: Le A 35 018
CURRENT APPLICATION NUMBER: US/09/988,863A
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 571
TYPE: DNA
ORGANISM: Pinus radiata
US-09-988-863A-5

Alignment Scores:

Pred. No.:	7,74e-05	Length:	571
Score:	124.50	Matches:	37
Percent Similarity:	48.21%	Conservative:	17
Best Local Similarity:	33.04%	Mismatches:	45
Query Match:	5.58%	Indels:	13
DB:	10	Gaps:	4

US-10-069-062-7 (1-432) x US-09-988-863A-5 (1-571)

QY 1 MetSerLyAlaPheSerAlaProGlyLySAlaPheLeuAlaGlyGlyTyrLeuValLeu 20
DB 244 ATGGCGTGTAGTGTGTGACCTCTGCTGTAAGTTTATATACAGAGCTTATCTATCTT 303
QY 21 GIUPROLLETyrAspAlaTyrValThrAlaLeuSerSerArgMetHisAlaValIlethr 40
DB 304 GAGAGCCAAATCCAGAGCTGTGCTTACACACAGCTGCTTACGCCATTTGGGAAG 363
QY 41 ProlYsglyThrSerLeuYsgLuserArg-----IleYsile 53
DB 364 CCACTCGGACTAGCAGATTCAGTACGTTGGGCAATGCTATGACAGATGGAATTA 423
QY 54 SerSerProGlnPheAlaAsnGlyGlyTyrGlyTyrHisIleSerSerAnthrGluYs 73
DB 424 ACATCCCTCAGCTTGCAGAG--GAGGCCATCTACAGCTATCTGTAGAGCTCTTAC 480
QY 74 ProArgGluVal-----GlnSerArgIleAsnProPheLeuGluAlaThrIlePhe 90
DB 481 CTGCAAAATGTTGCTTCTCAAGTACAAATGGAATCTTTGTGGAACAAGAGGCA 540
QY 91 IleValIleuAlaTyrIleGlnProthrGluAlaPhe 102
DB 541 TTGTGCTTGCAGCTGCAAAA-----GAAGCTTT 570

RESULT 36
US-09-918-740-18
Sequence 18, Application US/09918740
Publication No. US20030033626A1
GENERAL INFORMATION:
APPLICANT: Hahn, Frederick
APPLICANT: Kuehnle, Adelheid
TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
FILE REFERENCE: KAS-103XCI
CURRENT APPLICATION NUMBER: US/09/918,740
CURRENT FILING DATE: 2001-07-31

PRIOR APPLICATION NUMBER: 60/221,703

PRIOR FILING DATE: 2000-07-31

NUMBER OF SEQ ID NOS: 76

SOFTWARE: PatentIn version 3.0

SEQ ID NO 18

LENGTH: 1332

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Oligonucleotide containing N. tabacum and S. cerevisiae DNA

US-09-918-740-18

Alignment Scores:

Pred. No.:	0.000292	Length:	1332
Score:	124.50	Matches:	105
Percent Similarity:	37.34%	Conservative:	72
Best Local Similarity:	22.15%	Mismatches:	154
Query Match:	5.58%	Indels:	143
DB:	11	Gaps:	27

US-10-069-062-7 (1-432) x US-09-918-740-18 (1-1332)

QY 1 MetSerLyAlaPhe-----SerAlaProGlyLySAlaPheLeuAlaGlyGlyTyrLeu 18
DB 1 ATGTCATTACCGTCTTAACTTGTGCACCGGAAAGTTATTTTGTGGAACACTCT 60
QY 19 ValLeu-----GluProIleTyrAspAlaTyrValThrAlaLeuSerSerArgMetHis 36
DB 61 GCTGTGACACMACCCGCGCTGCTGTGTGTGTGCTGTT-----AGAACCTTAC 114
QY 37 AlaValIlethrProLyGlyThrSerLeuYsgLuserArgIleYsIleSerSerPro 56
DB 115 CTGCAATATAGC-----GAGTCATCGACACAGATCATTTGAATGTGACTTCCG 165
QY 57 GlnPheAla-----AsnGlyLutTyrGlyTyrHisIleSerSer----- 69
DB 166 GACATTACTTTATCATATAGTGTCCATCATATTTCAATGCCATCAGCAGATCAA 225
QY 70 ---AenThrGluYsProArgGluValGlnSerArgIleAsnProPheLeuGluAlaThr 88
DB 226 GTAACTCCCAAAATTTGGCAAGCTCA----- 255
QY 89 IlePheIleValIleuAlaTyrIleGlnProthrGluAlaPheAspLeuGluIleIle 108
DB 256 -----CAAGCACCGATGCTGTGTCCAGAACTGCTTACT 291
QY 109 TyrSerAspPro-----GlyTyrHisSerGlnIleAsp 119
DB 292 CTTTGGATCCGTTGTAGTCAACTATCCGATCTTCCACTTACCATGCA----- 342
QY 120 ThrGluThrYsThrSerSerAnGlyGlyYsThrPheLeuYr----- 134
DB 343 -----GCGTTTGTGTCTGATATATGTTTGTTCCTTA 375
QY 135 -----HisSerArgAlaIleThrGluValGlyYsThr-----Gly 146
DB 376 TGCCCCCATGCCAAGATATTAAGTTTCTTAAAGTCTTATACCATGCGTCTGG 435
QY 147 LeuGlySerSerAlaGlyLeuValSerValAlaIleThrSerLeuLeuSerHisPheIle 166
DB 436 TTGGGCTCAGCGCTCTATT-----TCTGTATCAGCGGCTTATGCTATGCGCTACTTG 489
QY 167 ProenValIleSerThrAsn-----LysAspIleLeuHisAsnValAlaGlnIleAla 184
DB 490 GGGGGCTTAAATGATCTTAATGACTTGGAAAGCTGTGCAAAAGATTAAGCATTTAGT 549
QY 185 HisCysTyrAla-----GlnYsIleYsIle-----GlySerGlyPheAspVal 198
DB 550 AATCAATGGGCTTCATAGTGAAGATATTCACGCTACCCCTTACGAGATATGATAC 609
QY 199 AlaThrAlaIleTyrGly-----LeuIleValYrThrGlyPheGlnProAlaLeuIle 216
DB 610 GCTGTGGCCACTTATGATATGCTGCTGCTATTTGAAAGACTCATATATGAAACAATA 669


```

Db      832 GAATGCCCTCAAGGCTTAGATCATGACTTAAGTAAGTAAGGACCGAT 891
Qy      309 ---GluysTyraSerSerProGluThrTyrr-----IleuSgluleuAphHisSer 325
Db      892 GACGAGGCTGTAAACCTAATATGAACGTATGAAACAACCTATGTAATGATTAAGATA 951
Qy      326 ValGluProLeuThrValAlaIleuysAsnIleArgIleSglYleuGlnAlaLeuThrGln 345
Db      952 AATCATGACGCTGCTTCTCAATCGGTGTTCTCAATCCCGATGATTAAGACTTATTAATAAT 1011
Qy      346 LysSerGluValProIleGluProAspValGlnThrGlnLeuLeuAspArgCysGlnGln 365
Db      1012 CTGAGCGGAT-----GATTGAGG-----GATTGAGG-----1029
Qy      366 IleProGlyCysValAlaGlyValAla---ValProGlyAlaGlyValTyraSphAlaIleAla 384
Db      1030 -----ATTGGCTCCACAAACCTTACCGGTGGTGGCGCGGCTGCTCTTGT 1077
Qy      385 ValLeu-----ValleuGluAenGlnValAlaGlnPheLysGlnLysThrLeu 400
Db      1078 ACTTGTTCAGAAAGACACTTACTCAAGACAAATTGACAGCTTCAAAAGAAATTGCA 1137
Qy      401 GluAenProAspTyraPheHisAsnValTyrrTpValAspLeu 414
Db      1138 GAT-----GATTGTAGTACGACGACATTTGAAACAGACTTG 1173

```

RESULT 38

```

US-10-069-909-4
; Sequence 4, Application US/1006909
; Publication No. US20030148479A1
; GENERAL INFORMATION:
; APPLICANT: KEASLING, JAY
; APPLICANT: MARTIN, VINCENT
; APPLICANT: PITTERA, DOUGLAS
; APPLICANT: KIM, SEON-WON
; APPLICANT: WITHERS III, SYDNOR T.
; APPLICANT: YOSHIKUNI, YASUO
; APPLICANT: NEWMAN, JACK
; APPLICANT: KHEBNIKOV, ARTEM VALENTINOVICH
; TITLE OF INVENTION: BIOSYNTHESIS OF ISOPENTENYL PYROPHOSPHATE
; FILE REFERENCE: 2000-0007
; CURRENT APPLICATION NUMBER: US/10/006,909
; CURRENT FILING DATE: 2002-04-02
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1332
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Mevalonate kinase nucleotide sequence
US-10-069-909-4

```

Alignment Scores:

```

Pred. No.: 0.000292 Length: 1332
Score: 124.50 Matches: 105
Percent Similarity: 37.348 Conservative: 72
Best Local Similarity: 22.154 Mismatches: 154
Query Match: 5.588 Indels: 143
DB: 12 Gaps: 27

```

US-10-069-062-7 (1-432) X US-10-069-909-4 (1-1332)

```

Qy      1 MetSerLysAlaPhe-----SerAlaProGlyLysAlaPheLeuAlaGlyLysThrLeu 18
Db      1 ARGTCATTCACCGTCTTACTTCTGCAACCGGAAAGGTATATTTTGTGACACCTCT 60
Qy      19 ValLeu-----GluProIleTyraSphAlaTyrrValThrAlaLeuSerSerArgMetHis 36
Db      61 GCTGTACACAAAGGCTGCGCTGCTGCTAGTGTCTGCGTGG-----AGAACCTTAC 114

```

```

Qy      37 AlaValIleThrProLysGlyThrSerLeuysGlnSerArgIleLysIleSerSerPro 56
Db      115 CTGCTAATTAAGC-----GAGTCATCTGACCAAGATACATTAATTTGACCTTCCG 165
Qy      57 GlnPheAla---AsnGlyGluThrGluTyrrHisIleSerSer-----69
Db      166 GACATTACCTTAAATCAATAGAGTGGTCATCATATGATTTCAATGCCATCACCGAGATCA 225
Qy      70 ---AsnThrGluLysProAspGluValAlaGlnSerArgIleAsnProPheLeuGlnAlaThr 88
Db      226 GTAACTCCCAAAAATTGGCCAAAGCTCA-----255
Qy      89 IlePheIleValLeuAlaTyrrIleGlnProThrGlnAlaPheAspLeuGluIleIle 108
Db      256 -----CAAGCCACCGATGCTGTCTCAGGAACCTCGTAGT 291
Qy      109 TyrSerAspPro-----GlyTyrrHisSerGlnLysP 119
Db      232 CTTTGGATCCGTTGTTAGCTCAACTATCCGAAATCTTCCACTACATCAGCA-----342
Qy      120 ThrGluThrLysThrSerSerAsnGlyLysThrPheLeuTyrr-----134
Db      343 -----CGCTTGTCTTCTGTATATAGTTGTTGTTGCCCTA 375
Qy      135 ---HisSerArgAlaIleThrGluValAlaGlyLysThr-----Gly 146
Db      376 TGCCCCCATGCCAATAATTAAGTTTCTTAAAGTCTTACTTTACCATGGGTGCTGG 435
Qy      147 LeuGlySerSerAlaGlyLeuValSerValAlaAlaThrSerLeuLeuSerHisPheIle 166
Db      436 TTGGGCTCAAGGCTCTTATTT-----TCTGTATCATCGGCTTGTGCTATAGCCATCTTG 489
Qy      167 ProAsnValIleSerThrAsn-----LysAspIleLeuHisAsnValAlaGlnIleAla 184
Db      490 GGGGCTTAATAGATCTTAATGACTTGAAGAACTGTGCAAGAAACATTAAGCATTAAGTG 549
Qy      185 HisCysTyraAla-----GlnLysLysIle-----GlySerGlyPheAspVal 198
Db      550 AATCAATGGGCTTCTATAGTGAAGAAAGTATTTACCGGTACCCCTTCAAGATATATTAAC 609
Qy      199 AlaThrAlaIleTyrrGly-----LeuIleValTyrrArgArgPheGlnProAlaLeuIle 216
Db      610 GCTGTGGCCACTTATGATATGCCCTGCTATTTGAAAAAGACTCAATATGAAGCAATA 669
Qy      217 Asn---AspValPheGlnValLeuGlnLysSerAspProGlyLysPheProThrGluLeuLys 235
Db      670 AACACAAACATTTTAAGTCTTATGATGATTTCCA-----705
Qy      236 LysLeuIleGlnSerAsnTrpGlnGluLysHisGluArgCysThrLeuProTyrrGlyIle 255
Db      706 -----GCCAAT 711
Qy      256 LysLeuLeuMetGlyAspValLysGlyLysGlySerGluThrProLysLeuValSerArgVal 275
Db      712 CCAATGATCTTAACCTAATACCTAATGCAATTCGCAAGCTCAAAAGAAATCTTGTGCTGCGCT 771
Qy      276 LeuGlnTrpLysGluLys---ProGlnGlnLysSerValValTyraSglLeuLeuAsn 294
Db      772 CGTGTGTGGTGCACCGAAGAAATTTCTTAAGTTATGAAGCAATTCATGATGCCATGGCT 831
Qy      295 SerAlaAsnLeuGln-----PheMetLysGlnLeuArgGluMetArg-----308
Db      832 GAATGTGCCCTTCAAGGCTTAGAGATCATGACTAAGTTAAGTAATGAAGGACCGAT 891
Qy      309 ---GluysTyraSerSerProGluThrTyrr-----IleuSgluleuAphHisSer 325
Db      892 GACGAGGCTGTAAACCTAATATGAACGTATGAAACAACCTATGTAATGATTAAGATA 951
Qy      326 ValGluProLeuThrValAlaIleuysAsnIleArgIleSglYleuGlnAlaLeuThrGln 345
Db      952 AATCATGACGCTGCTTCTCAATCGGTGTTCTCAATCCCGATGATTAAGACTTATTAATAAT 1011
Qy      346 LysSerGluValProIleGluProAspValGlnThrGlnLeuLeuAspArgCysGlnGln 365

```

```

Db      1012 CTGACGAT-----GATTTGACA-----1029
Qy      366 ILEProGlyCyValGlyVal---ValProGlyAlaGlyGlyTyrAspAlaIleAla 384
Db      1030 -----ATTGGCTCCCAAAACTTACCGGGTGGGGCGGCTTCTTTG 1077
Qy      385 ValLeu-----ValLeuGluGlnGlnValGlyAsnPhelyGlnLysThrLeu 400
Db      1078 ACTTTGTTACGAAGACATTACTCAAGGCAATTGACAGCTTCAAAAGAAATTGCAA 1137
Qy      401 GluAsnProAspTyrPheHisAsnValTyrTrpValAspLeu 414
Db      1138 GAT-----GATTTAGTTACGACGACATTGAAAGACACTTG 1173

RESULT 39
US-09-938-842A-1231
; Sequence 1231, Application US/0993842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1231
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1231

Alignment Scores:
Pred. No.: 0.00034 Length: 1137
Score: 123.00 Matches: 94
Percent Similarity: 128.00% Conservat: 77
Best Local Similarity: 20.89% Mismatches: 149
Query Match: 5.52% Indels: 132
          10 Gaps: 17

US-10-069-062-7 (1-432) x US-09-938-842A-1231 (1-1137)
Qy      7 AlaProGlyValAlaPheLeuAlaGlyGlyTyrLeuValLeuGluProIleTyrAspAla 26
Db      19 GCTCTGGGAAGATCATCTTCGACGGGAAACGCTGTT-----GTTTCATGATGCC 69
Qy      27 TyrValThrAlaLeuSerSerArgMetHisAlaValIleThr-----ProIleGly 43
Db      70 ACCGCTGATAGCTCCGCCATTCATCTTACACTTACGTTACTCTCCGCTTCTCTTCCA 129
Qy      44 ThrSerLeuLysGluSerArgIleLysIle-SerSerProGlnPheAlaAsnGlyLysTr 63
Db      130 TCAGCTGGAACAAATGATAGAGCTTACCTCAGCTCAAGGACATTT-----CCTTG 180
Qy      63 PGIuLysIleLysSerSerAsnThrGluLysProArgGluValGlnSerArgIleAsnPr 83
Db      181 GAGTTTATGCG-TCC-----195
Qy      83 CPhenLeuGluAlaThrIlePheIleValLeuAlaTyrIleGlnProThrGluAlaPheAs 103
Db      196 -----TTAGCCAGAAATCAAAAGACGATTCCTTATGA 227
Qy      103 PLeuGluIleIleIleTyrSerAspProGlyTyrHisSerGlnGlu-----118

```

```

Db      228 TTCAAGACATCTCTCCGCTTACAGCGGCTTCATGTTCAAGAGACCTTAATCAAT 287
Qy      119 -----AspThrGlnThrLysThrSerSerAsnGly 129
Db      288 TGCAGTTTGGTTGAAGACAAATCTTCCAAAGGAAAGATGGCTCTCTCTGGGAT 347
Qy      129 UlySerPheLeuTyr---HisSerArgAlaIle-----139
Db      348 CTCACGTTTCTCTGTTATATACCAAGAAATTATAGGTTCAATCCGCTACAGTATCAT 407
Qy      140 ---ThrGluValGlu---LysThrGlyLeuGlySerSerArgIleGlyValLeuValSerVal 157
Db      408 TAACCTAGCTTCATACAGGCTCGGCTCGGCTTCATACAGAGCTTATGTAGTCTCT 467
Qy      157 LAlaThrSerLeuSerHisPheIle-----ProAs 168
Db      468 CACAGCTGCTCTCTGCTTCTCTTATTTACAGAGAAACCGGTGTAAAGGTTGTCATC 527
Qy      168 nValIleSerThrAsnLysAspIleLeuHisAsnValAlaGlnIleAlaHisCysTyrAl 188
Db      528 TCTCATTAACCAATCTTGAAGTCTTAATTAATAGGCT-----TTGCA 572
Qy      188 aGlnLysLysIle-----GlySerGlyPheAspValAlaThrAlaIleTyrG 204
Db      573 AGCGGAAAGATCATCATCAGGAAACCTTGGATACAAACCCGTCAGTACATACG 632
Qy      204 YLeuIleValTyrArgArgPheGlnProAlaLeuIleAsnAspValPheGlnValLeu 224
Db      633 CAACATGATC-----642
Qy      224 uSerAspProGluLysPheProThrGlnLeuLysLysLeuIleGluSerAsnTrpGlu 244
Db      643 -----AGTTCTGCTCAGCGAATAATCTCGTTACATCAAC-----681
Qy      244 uLysHisGluArgCysThrLeuProTyrGlyIleLysLeuLeuMetGlyAspValLysG 264
Db      682 -----ATGCGCT-----CTGAAATGTTAATTACCAACATCAGAGT 716
Qy      264 YGlySerGluThrProLysLeuValSerArgValLeuGlnTrpLysGluLysProG 284
Db      717 TGGCGAAACAAAGCTCTGCTCTGCTGTGTCTCAGAGAGCGGTAAACATCTCGA 776
Qy      284 uGluSerSerValValTyrAspGlnLeuAsnSerAlaAsnLeuGlnPheMetLysGlu 304
Db      777 TGCATGAAGCATGTTTCAACGCCGTGATTTCTATAGCAAAAGCTCGCTGCATCAT 836
Qy      304 uArgGluMetArgGluLysTyrAspSerAspProGluThrTyrIleLysGluLeuAsp 324
Db      837 TCAGCTTAAGACGACCTCAGTTACAGAAAGAAAGAGAGAAATTAAGAAATCATGGA 896
Qy      324 sSerValGluProLeuThrValAlaIleLysAsnIleArgLysGlyLeuGlnAlaLeu 344
Db      897 GATGAACCAAGCTGTCTCTCTCAATGGGGTTAGCCACAGCTCAATCGAGGCTGT 954
Qy      344 rGlnLysSerGluValProIleGluProAspValGlnThrGlnLeuLeuAspArgCys 364
Db      955 -----ATTTAACCAAGCTGTAAGACAAAGCTTGTCTCAAA-----990
Qy      364 nGluIleProGlyCyValGlyValGlyValProGlyAlaGlyGlyTyrAspAlaIle 384
Db      991 -----CTTACAGAGAGCTGTGGCGGCGGCTGTCT 1022
Qy      384 aValLeuValLeuGlnGlnValGlyAsnPhelyGlnLysThrLeuGluAsnProAs 404
Db      1023 CACTCTATTACCAACCGGACGGTGGTGACAAAGTGATGAGAGAGCTCGAGTCAGCGG 1082
Qy      404 pTyr-----PheHisAsnValTyrTrp 411
Db      1083 -TTTCACTGTTTACAGGCAATGATGG 1109

RESULT 40
US-09-815-242-9572

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2003, 09:41:12 ; Search time 23 Seconds

(without alignments)
883.284 Million cell updates/sec

Title: US-10-069-062-7

Perfect score: 432

Sequence: 1 MSKAFSAPGKAFIAGLYLV.....DLEQTEGVLEEKEDYIGL 432

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	2.5	451	1 ERG8_YEAST	P24521 saccharomyc
2	8	1.9	224	1 R101_ORYSA	P45635 oryza sativ
3	8	1.9	524	1 THS3_HALVO	Q9phaz halobacteri
4	7	1.6	99	1 Y073_NPVAC	P41472 autographa
5	7	1.6	107	1 Y065_MYCTU	P95094 mycobacteri
6	7	1.6	108	1 U136_ARATH	O64847 arabidopsis
7	7	1.6	108	1 YF33_PSEAB	O91310 pseudomonas
8	7	1.6	133	1 R122_AQUAB	O64436 aquiflex ae
9	7	1.6	133	1 R122_AQUAB	O92145 aquiflex pyr
10	7	1.6	160	1 PHNA_SPIPL	P72504 spirulina p
11	7	1.6	170	1 R1M1_XANNC	Q8bmy2 xanthomonas
12	7	1.6	179	1 Y140_METUA	Q9h6k4 homo sapien
13	7	1.6	206	1 OPA3_HUMAN	Q57605 methanococ
14	7	1.6	217	1 ELBB_ECOLI	P26428 escherichia
15	7	1.6	219	1 R103_ORYSA	Q40649 oryza sativ
16	7	1.6	220	1 R110_MAIZE	P45633 zea mays (m
17	7	1.6	228	1 CDGA_RICPR	O92548 r phosphati
18	7	1.6	264	1 KKA3_ENTRA	P00554 enterococcu
19	7	1.6	265	1 PRAP_SALTY	Q06995 salmonella
20	7	1.6	266	1 P1RC_KLEBN	O91480 klebsiella
21	7	1.6	291	1 HCHA_PSEAB	O91480 pseudomonas
22	7	1.6	298	1 NADA_FUSNN	Q846c9 fusobacteri
23	7	1.6	341	1 FLAA_BORUN	P70856 borrelia bu
24	7	1.6	348	1 UN97_CAEBL	P50464 caenorhabdi
25	7	1.6	396	1 SOTB_HABEN	P44535 haemophilus
26	7	1.6	401	1 ASSY_XYLEA	Q9pemy xyella fas
27	7	1.6	404	1 ASSY_XYLEA	P55606 xyella fas
28	7	1.6	404	1 KIME_SCHPO	Q09780 schizosacch
29	7	1.6	407	1 EYV1_HUMAN	P49640 homo sapien
30	7	1.6	420	1 PSD4_SCHMA	O17453 schistosoma
31	7	1.6	431	1 YG84_METTH	O27719 methanobact
32	7	1.6	496	1 AMPA_HELPJ	Q921r1 helicobacte
33	7	1.6	496	1 AMPA_HELPJ	O25294 helicobacte

34	7	1.6	517	1 CPT7_ORYIA	P70085 oryzias lat
35	7	1.6	542	1 CITS_BACCU	O34427 bacillus su
36	7	1.6	628	1 YE1F_SCHPO	O13879 schizosacch
37	7	1.6	631	1 IL16_HUMAN	O14005 homo sapien
38	7	1.6	664	1 YRGK_CAEBL	O18696 caenorhabdi
39	7	1.6	747	1 ELS_BOVIN	P04985 bos taurus
40	7	1.6	750	1 ELS_CHICK	P07916 gallus gall
41	7	1.6	833	1 HMDH_AGRIP	O76819 agrotis ips
42	7	1.6	838	1 YRAJ_ECOLI	P42915 escherichia
43	7	1.6	897	1 SAP1_YEAST	P39955 saccharomyc
44	7	1.6	1037	1 ACRD_ECOLI	P24177 escherichia
45	7	1.6	1087	1 XYNX_CLOTM	P38535 clostridium

ALIGNMENTS

```

RESULT 1
ERG8_YEAST      STANDARD;      PRT;      451 AA.
AC      P24521:
DT      01-MAR-1992 (Rel. 21, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      Phosphomevalonate kinase (EC 2.7.4.2).
GN      ERG8 OR YMR220W OR YM9959.02.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE OF 1-424 FROM N.A.
RX      MEDLINE=91117228; PubMed=1846667;
RA      Tsay Y.H., Robinson G.W.;
RA      "Cloning and characterization of ERG8, an essential gene of
RT      Saccharomyces cerevisiae that encodes phosphomevalonate kinase.";
RL      Mol. Cell. Biol. 11:620-631(1991).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=5288C / AB972;
RX      PubMed=9169872;
RA      Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA      Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA      Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA      Rice P., Skellton J., Walsh S., Whitehead S., Barrett B.G.;
RT      "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT      XII."
RL      Nature 387:90-93(1997).
CC      -1- FUNCTION: ERG8 IS AN ESSENTIAL GENE.
CC      -1- CATALYTIC ACTIVITY: ATP + (R)-5-phosphomevalonate = ADP + (R)-5-
CC      -1- dihydrophosphomevalonate.
CC      -1- PATHWAY: SECOND STEP IN ISOPRENTENYL DIPHOSPHATE FORMATION.
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC      -1- SIMILARITY: BELONGS TO THE GMP KINASE FAMILY. MEVALONATE KINASE
CC      SUBFAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
DR      EMBL; M63648; AA34596.1; -
DR      EMBL; Z49939; CA90191.1; -
DR      PIR; S57588; S57588.
DR      SGD; S0004833; ERG8.
DR      InterPro; IPR006203; GHMPkinse_ATP.
DR      InterPro; IPR005916; Pmev_kin_erg8.
DR      Pfam; PF00288; GHMP_kinases; 1.
DR      TIGRFAMs; TIGR01219; Pmev_kin_ERG8; 1.
DR      PROSITE; PS00627; GHMP_KINASES_ATP; 1.

```

KW Transferase; Kinase; ATP-binding; Isoprene biosynthesis;
 KW Sterol biosynthesis.
 FT NP BIND 150 160 ATP (POTENTIAL).
 FT BINDING 183 183 ATP (POTENTIAL).
 FT CONFLICT 213 213 A -> R (IN REF. 1).
 FT CONFLICT 418 423 TANDKR -> PLMTKD (IN REF. 1).
 SQ SEQUENCE 451 AA; 50455 MW; DB2B6C862153683C CRC64;
 Query Match
 Best Local Similarity 100.0%; Pred. No. 0.0045; Length 451;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 144 KTGLSSAGLV 154
 DB 152 KTGLSSAGLV 162
 RESULT 2
 ID R101 ORYSA STANDARD; PRT; 224 AA.
 AC P45635;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DB 60S ribosomal protein L10-1 (Putative tumor suppressor SC34).
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriactoidae; Oryzae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Indica-IR36;
 RA Kim J.K.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE L10E FAMILY OF RIBOSOMAL PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X81691; CAA57339.1; -.
 DR PIR; S49575; S49575.
 DR GRAMENE; P45635; -.
 DR InterPro; IPR001197; Ribosomal_L10E.
 DR Pfam; PF00826; Ribosomal_L10E; 1.
 DR TIGRFAMs; TIGR00279; L10E; 1.
 DR PROSITE; PS01257; RIBOSOMAL_L10E; 1.
 KW Ribosomal protein.
 SQ SEQUENCE 224 AA; 25317 MW; 1FBD74224BED029 CRC64;
 Query Match
 Best Local Similarity 1.9%; Score 8; DB 1; Length 224;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 AFGKAFIA 14
 DB 210 AFGKAFIA 217
 RESULT 3
 ID THS3 HALVO STANDARD; PRT; 524 AA.
 AC Q9HHA2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thermosome subunit 3 (Heat shock protein CCT3).
 CC CCT3.

OS Halobacterium volcanii (Haloflex volcanii).
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Haloflex.
 OX NCBI_TaxID=2246;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DS70;
 RA Kovacs E., Lund P.A.;
 RT "Sequence of the cct3 gene, a third chaperonin homolog in the
 RT halophilic archaeon Haloflex volcanii."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MOLECULAR CHAPERONE; BINDS UNFOLDED POLYPEPTIDES IN
 CC VITRO, AND HAS A WEAK ATPASE ACTIVITY (BY SIMILARITY).
 CC -1- SUBUNIT: FORMS AN OLIGOMERIC COMPLEX OF EIGHT-MEMBERED RINGS
 CC (BY SIMILARITY).
 CC -1- INDUCTION: By heat shock.
 CC -1- SIMILARITY: Belongs to the TCP-1 chaperonin family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF298660; AAG17906.1; -.
 DR HSSP; P48424; 1A6D.
 DR InterPro; IPR002194; Chaperonin_TCP-1.
 DR InterPro; IPR001844; Chaperin_Cpn60.
 DR InterPro; IPR002423; Cpn60_TCP-1.
 DR Pfam; PF00118; Cpn60_TCP1; 1.
 DR PRINTS; PR00298; CHAPERONIN60.
 DR PRINTS; PR00304; TCOMPLEXTCP1.
 DR PROSITE; PS00750; TCP1_1; 1.
 DR PROSITE; PS00751; TCP1_2; 1.
 DR PROSITE; PS00995; TCP1_3; 1.
 KW Chaperone; ATP-binding; Heat shock.
 SQ SEQUENCE 524 AA; 55261 MW; CB81E2EAF91181B9 CRC64;
 Query Match
 Best Local Similarity 1.9%; Score 8; DB 1; Length 524;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 371 GGVPGAG 378
 DB 403 GGVPGAG 410
 RESULT 4
 ID Y073 NPVAC STANDARD; PRT; 99 AA.
 AC P41472;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 11.5 kDa protein in IAP2-VLP1 intergenic region.
 OS Autographa californica nuclear polyhedrosis virus (AcNPV).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 OC Nucleopolyhedrovirus.
 OX NCBI_TaxID=46015;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C6;
 RX MEDLINE=94303173; PubMed=8030224;
 RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
 RT "The complete DNA sequence of Autographa californica nuclear
 RT polyhedrosis virus."
 RL Virology 202:586-605(1994).
 CC -1- SIMILARITY: TO CORRESPONDING ORF IN OPNPV.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL, I22858, AAA66703.1; -
DR PIR, B72859, B72859.
KW Hypothetical protein.
SQ SEQUENCE 99 AA; 11526 MW; C65481B8A76A0EA4 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 99;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 KKLIESN 241
DB 57 KKLIESN 63

RESULT 5

YU65 MYCTU STANDARD; PRT; 107 AA.
AC P95094;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein RV3065.
GN RV3065 OR MTJ3150.1 OR MTCY22D7.17C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gae S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsbury T., Ungles K., Krogan A., McLean U., Moulé S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence";
RL Nature 393:537-544 (1998).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Bisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SMALL MULTIDRUG RESISTANCE (SMR)
CC PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL, Z83866, CA806258.1; -
DR EMBL, AE007132, AAK47485.1; -

DR PIR, B70650, B70650.

DR TIGR, MT3150.1; -
DR TubercuList; RV3065; -
DR InterPro; IPR000390; Smr.

DR Pfam; PF00893; Multi Drug Res; 1.

KW Hypothetical protein; Transmembrane; Transport; Complete proteome.

FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 30 50 POTENTIAL.
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 79 99 POTENTIAL.
SQ SEQUENCE 107 AA; 11083 MW; 3D372AC2FAE89834 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 VVATSL 162
DB 14 VVATSL 20

RESULT 6

U136 ARATH STANDARD; PRT; 108 AA.
ID U136 ARATH
AC O64847;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein At2g26240.
GN AT2G26240 OR T1D16.12.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids; II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman J.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nieman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana";
RL Nature 402:761-768 (1999).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0136 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL, AC004484; AAC14534.1; -
DR PIR, A84658; A84658.
DR InterPro; IPR005349; UPF0136.
DR Pfam; PF03647; UPF0136; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 5 25 POTENTIAL.
FT TRANSMEM 85 105 POTENTIAL.
SQ SEQUENCE 108 AA; 11108 MW; 7288CD81200763C0 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 AGLVSVV 157
 DB 86 AGLVSVV 92

RESULT 7
 YF33_PSEAE STANDARD; PRT; 108 AA.
 ID YF33_PSEAE 091310;
 AC 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hypothetical UPF0133 protein PA1533.
 GN PA1533;
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PAOI;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Watterer P.,
 RA Hickey M.J., Brinkman P.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Mong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V., an
 RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
 RT opportunistic pathogen.";
 RT Nature 406:959-964(2000).
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SIMILARITY: Belongs to the UPF0133 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AE004581; AAG04922.1; -;
 DR PIR; B83455; B83455.
 DR HAMAP; MF 00274; -; 1.
 DR InterPro; IPR004401; Cons_hypoch103.
 DR Pfam; PF02575; DUF149; 1.
 DR TIGRPFAM; TIGR00103; TIGR00103; 1.
 DR TIGRPFAM; TIGR00103; TIGR00103; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 108 AA; 11877 MW; 6E0E82F71BEE99D7 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 108;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 AGLVSVV 157
 DB 38 AGLVSVV 44

RESULT 8
 RL22_AQUAE STANDARD; PRT; 133 AA.
 ID RL22_AQUAE 066436;
 AC 066436;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 50S ribosomal protein L22.
 GN RPLV OR AQ_016 OR AQ_016a.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
 OX NCBI_TaxID=63363;

RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF5;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young M.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Shear M.A., Keller M., Aubay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus.";
 RT Nature 392:353-358(1998).
 RN [2]
 RP CONCEPTUAL TRANSLATION.
 RA Veuchey A.-L.;
 RL Unpublished observations (MAR-2000).
 CC -1- FUNCTION: THIS PROTEIN BINDS SPECIFICALLY TO 23S RNA; ITS BINDING
 CC IS STIMULATED BY OTHER RIBOSOMAL PROTEINS, E.G. L4, L17, AND L20.
 CC IT IS IMPORTANT DURING THE EARLY STAGES OF 50S RECONSTITUTION
 CC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE L22P FAMILY OF RIBOSOMAL PROTEINS.
 CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. TWO FRAMESHIFTS WERE
 CC INTRODUCED IN POSITIONS 30 AND 54 TO PRODUCE THIS ORF.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AE000669; AAC06396.1; ALT_FRAME.
 DR HSSP; P48286; 1BXE.
 DR InterPro; IPR005727; L22 bact org.
 DR InterPro; IPR001063; Ribosomal_L22.
 DR Pfam; PF00237; Ribosomal_L22; 1.
 DR ProDom; PD001032; Ribosomal_L22; 1.
 DR TIGRPFAM; TIGR01044; rplV_bact; 1.
 DR PROSITE; PS00464; RIBOSOMAL_L22; 1.
 KW Ribosomal protein; RNA-binding; Complete proteome.
 SQ SEQUENCE 133 AA; 15013 MW; C46590A7313DF97 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 133;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 VLEEKPE 427
 DB 122 VLEEKPE 128

RESULT 9
 RL22_AQUAPY STANDARD; PRT; 133 AA.
 ID RL22_AQUAPY 092145;
 AC 092145;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 50S ribosomal protein L22.
 GN RPLV OR RPL22.
 OS Aquifex pyrophilus.
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
 OX NCBI_TaxID=2714;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=KO15A / DSM 6858;
 RX MEDLINE=20254845; PubMed=10795828;
 RA Bocchetta M., Gribaldo S., Sanangelantoni A.M., Cammarano P.,
 RA "Phylogenetic depth of the bacterial genera Aquifex and Thermotoga
 RT inferred from analysis of ribosomal protein, elongation factor, and
 RT RNA polymerase subunit sequences.";
 RL J. Mol. Evol. 50:366-380(2000).
 CC -1- FUNCTION: THIS PROTEIN BINDS SPECIFICALLY TO 23S RNA; ITS BINDING
 CC IS STIMULATED BY OTHER RIBOSOMAL PROTEINS, E.G. L4, L17, AND L20.

IT IS IMPORTANT DURING THE EARLY STAGES OF 50S RECONSTITUTION
(BY SIMILARITY).
-1- SIMILARITY: BELONGS TO THE L22P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF040100; AAD08790.1; -
DR HSP; P48286; 1BXE.
DR InterPro; IPR005727; L22 bact org.
DR InterPro; IPR001063; Ribosomal L22.
DR Pfam; PF00237; Ribosomal_L22; 1.
DR ProDom; PD001032; Ribosomal_L22; 1.
DR TIGRPFams; TIGR01044; rplv Bact; 1.
DR PROSITE; PS00464; RIBOSOMAL_L22; 1.
KW Ribosomal protein; rRNA-binding.
SQ SEQUENCE 133 AA; 15118 MW; 894D0ED96A604247 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 133;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 VLEEKPE 427
Db 122 VLEEKPE 128

RESULT 10
ID PHA_SPIPL STANDARD; PRT; 160 AA.
AC P72504; O08136;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Allophycocyanin alpha chain.
GN APCA.
OS Spirulina platensis.
OC Bacteria; Cyanobacteria; Oscillatoriales; Spirulina.
OX NCBI_Taxid=1156;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Italy;
RA Meesapodsuk D., Chetkul W., Nomsasawai P., Anjard C.,
RA Tanticharoen W., Chevananarak S.;
RT "Organization and nucleotide sequence of the a, b and c subunits of
RT allophycocyanin genes from Spirulina platensis";
RU Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=F3;
RA Qin S., Kawata Y., Yano S., Tseng C., Kojima H.;
RT "Cloning and sequencing of the allophycocyanin genes from
RT cyanobacterium Spirulina platensis";
RU Submitted (MAY-1997) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: LIGHT-HARVESTING PHOTOSYNTHETIC BILE PIGMENT-PROTEIN
CC FROM THE PHYCOBILIPROTEIN COMPLEX. ALLOPHYCOCYANIN HAS A MAXIMUM
CC ABSORPTION AT APPROXIMATELY 650 NANOMETERS.
CC -1- SUBUNIT: Heterodimer of an alpha and a beta chain.
CC -1- PTM: CONTAINS ONE COVALENTLY LINKED BILIN CHROMOPHORE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X95898; CA65141.1; -
DR EMBL; D86179; BAA19985.1; -
DR PIR; S55557; S55557.
DR PDB; 1ALL; 11-JUN-96.
DR InterPro; IPR001659; Phycobilisome.
DR Pfam; PF00502; Phycobilisome; 1.
DR ProDom; PD000340; Phycobilisome; 1.
KW Phycobilisome; Electron transport; Photosynthesis; Bile pigment;
KW 3D-structure.
FT INIT MET 0 0 BY SIMILARITY.
FT BINDING 80 80 PHYCOCYANOBILIN CHROMOPHORE.
SQ SEQUENCE 160 AA; 17260 MW; FA6FC8ADFAD0EFA4 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 160;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 VATSILS 163
Db 135 VATSILS 141

RESULT 11
ID RIMM_XANAC STANDARD; PRT; 170 AA.
AC RIMM_XANAC
OS O8PWY2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable 16S rRNA processing protein rlmW.
GN RIMM OR XAC1293.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_Taxid=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
RA Camarotte G., Camaravan F., Cardozo J., Chambergro F., Clapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorxy H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locall E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spicola L.A.F., Takita M.A., Tamara R.E., Teixeira B.C., Tezza R.I.D.,
RA Trindade dos Santos M., Trufi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities";
RU Nature 417:459-463(2002).
CC -1- FUNCTION: Essential for efficient processing of 16S rRNA. Probably
CC part of the 30S subunit prior to or during the final step in the
CC processing of 16S free 30S ribosomal subunits. It could be some
CC accessory protein needed for efficient assembly of the 30S
CC subunit. It is needed in a step prior to rbfA during the
CC maturation of 16S rRNA. It has affinity for free ribosomal 30S
CC subunits but not for 70S ribosomes (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC -1- SIMILARITY: BELONGS TO THE RIMM FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC or send an email to license@isb-sib.ch.

 CC EMBL, AF011759; AM36164.1; -.
 CC DR HAMAP, MF_00014; 1.
 CC DR Interpro, IPR002676; RIMM.
 CC DR Pfam, PF05239; PRC; 1.
 CC DR Pfam, PF01782; RIMM; 1.
 CC RNA processing; Complete proteome.
 CC KW RRNA processing; Complete proteome.
 CC SQ SEQUENCE 170 AA, 19195 MW, 7299466803C5228 CRC64;

 CC Query Match 1.6%; Score 7; DB 1; Length 170;
 CC Best Local Similarity 100.0%; Pred. No. 25;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 CC Db 410 YWVDLEE 416
 CC |||||
 CC 102 YWVDLEE 108

 CC RESULT 12
 CC OPA3 HUMAN STANDARD; PRT; 179 AA.
 CC AC Q9H6K4;
 CC DT 28-FEB-2003 (Rel. 41, Created)
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
 CC DE Optic atrophy 3 protein.
 CC GN OPA3.
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC OX NCBI_Taxid=9606;
 CC RN [1]
 CC RP SEQUENCE FROM N.A., DISEASE, AND TISSUE SPECIFICITY.
 CC RA MEDLINE=21559913; PubMed=1168429;
 CC RA Anilster Y., Kieta R., Shaag A., Gahl W.A., Elpeleg O.,
 CC RT "Type III 3-methylglutaconic aciduria (optic atrophy plus syndrome, or
 CC RT Costeff optic atrophy syndrome): identification of the OPA3 gene and
 CC RT its founder mutation in Iraqi Jews.";
 CC RL Am. J. Hum. Genet. 69:1218-1224(2001).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
 CC RA Suzuki Y., Ohtsushi M., Nishi T., Shibahara T., Tanaka T.,
 CC RA Nakamura Y., Isegai T., Sugano S.,
 CC RT "NEO human cDNA sequencing project.";
 CC RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 CC RN [3]
 CC RP SEQUENCE FROM N.A.
 CC RC TISSUE=Kidney;
 CC RX MEDLINE=22388257; PubMed=12477932;
 CC RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 CC RA Klausner R.D., Collins F.S., Wagner L., Schmeien C.M., Schuler G.D.,
 CC RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
 CC RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 CC RA Diachenko L., Marushina K., Farmer A.A., Rubin G.M., Hong L.,
 CC RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Schetz T.E.,
 CC RA Brownstein M.J., Usdin T.B., Toshjuki S., Carninci P., Prange C.,
 CC RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 CC RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 CC RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 CC RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 CC RA Fahy J., Helton E., Kerteman M., Madan A., Rodrigues S., Sanchez A.,
 CC RA Whiting M., Madan A., Young A.C., Shcherchenko Y., Bouffard G.G.,
 CC RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 CC RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 CC RA Buterfield J.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
 CC RA Scherch A., Schein J.E., Jones S.J.M., Maria M.A.,
 CC RT "Generation and initial analysis of more than 15,000 full-length
 CC RT human and mouse cDNA sequences.";
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC CC -1- FUNCTION: May play some role in mitochondrial processes.
 CC CC -1- TISSUE SPECIFICITY: Ubiquitous. Most prominent expression in

CC skeletal muscle and kidney.
 CC -1- DISEASE: Defects in OPA3 are the cause of type III 3-
 CC methylglutaconic aciduria (MGA type III); also known as optic
 CC atrophy plus syndrome or Costeff optic atrophy syndrome. MGA type
 CC III is a neuroophthalmologic syndrome consisting of early-onset
 CC bilateral optic atrophy and later-onset spasticity, extrapyramidal
 CC dysfunction, and cognitive deficit. Urinary excretion of 3-
 CC methylglutaconic acid and of 3-methylglutaric acid is increased.
 CC -1- SIMILARITY: BELONGS TO THE OPA3 FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC DR EMBL, AK025840; BAB15255.1; -.
 CC DR EMBL, BC005059; AAH05059.1; -.
 CC DR Genew; HGNC:8142; OPA3.
 CC DR MIM; 606580; -.
 CC DR MIM; 258501; -.
 CC KW Mitochondrion; Coiled coil; Vision.
 CC FT DOMAIN 103 163 COILED COIL (POTENTIAL).
 CC SQ SEQUENCE 179 AA, 19996 MW, 2F848F84B24E5A8 CRC64;

 CC Query Match 1.6%; Score 7; DB 1; Length 179;
 CC Best Local Similarity 100.0%; Pred. No. 26;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 CC Db 86 EATIFIV 92
 CC |||||
 CC 86 EATIFIV 92

 CC RESULT 13
 CC Y140 META STANDARD; PRT; 206 AA.
 CC ID Y140 META
 CC AC 057605;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Hypothetical protein M0140.
 CC GN M0140.
 CC OS Methanococcus jannaschii.
 CC OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 CC OC Methanocaldococcaceae; Methanocaldococcus.
 CC OX NCBI_Taxid=2190;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 CC RX MEDLINE=96337999; PubMed=8688087;
 CC RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 CC RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 CC RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 CC RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 CC RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Puhmann J.L., Nguyen D.,
 CC RA Utechtack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 CC RA Cotton M.P., Roberts K.M., Huzar M.A., Kaine B.P., Borodovsky M.,
 CC RA Klenk H.-P., Frazer C.M., Smith H.O., Weese C.R., Venter J.C.,
 CC RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 CC RT jannaschii.";
 CC Science 273:1058-1073(1996).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U67471; AAB98123.1; --
 DR FIR; E64317; E64317.
 DR TIGR; M0140; --
 DR InterPro; IPR006367; Cysg N.
 DR TIGR/Pfam; TIGR01470; Cysg Nterm; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 206 AA; 23989 MM; 50C36E48E4FBC44 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 206;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ELKRLIE 239
 |||||
 DB 193 ELKRLIE 199

RESULT 14
 EMBL ECOLI STANDARD; PRT; 217 AA.
 AC P26428; P76673;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Enhancing lycopene biosynthesis protein 2 (Sigma cross-reacting
 protein 27A) (SCR-27A).
 GN EMBL OR EMBL OR B3209 OR SF3249.
 OS Escherichia coli, and
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562; 623;
 [1]
 RN SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=K12 / W3110;
 RA Smilie D.A., Fujita N., Townsley F.M., Ishihama A., Hayward R.S.;
 RT "Sequence and characterisation of the gene encoding the sigma
 cross-reacting protein 27A in Escherichia coli.";
 RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=K12 / W61655;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474 (1997).
 [3]
 RN SEQUENCE OF 1-24.
 RC SPECIES=E.coli;
 RX MEDLINE=92246944; PubMed=1575737;
 RA Ueshima R., Fujita N., Ishihama A.;
 RT "Identification of Escherichia coli proteins cross-reacting with
 antibodies against region 2.2 peptide of RNA polymerase sigma
 subunit.";
 RL Biochem. Biophys. Res. Commun. 184:634-639 (1992).
 [4]
 RN SEQUENCE FROM N.A.
 RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
 RX MEDLINE=22272406; PubMed=12384590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Sun L., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
 through comparison with genomes of Escherichia coli K12 and O157.";
 RL Nucleic Acids Res. 30:4432-4441 (2002).
 [5]
 RN POSSIBLE FUNCTION.
 RC SPECIES=E.coli;

RX MEDLINE=98269038; PubMed=9603997;
 RA Hemmi H., Ohnuma S., Nagaoka K., Nishino T.;
 RT "Identification of genes affecting lycopene formation in Escherichia
 coli transformed with carotenoid biosynthetic genes: candidates for
 early genes in isoprenoid biosynthesis.";
 RL J. Biochem. 133:1088-1096 (1998).
 CC -1- FUNCTION: MAY BE INVOLVED IN THE EARLY STEPS OF ISOPRENOID
 BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE ESI FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; D13188; BAA02487.1; --
 DR EMBL; U18997; AAB58011.1; ALT INIT.
 DR EMBL; AE000400; AAC76241.1; ALT INIT.
 DR EMBL; AE015335; AAN44714.1; ALT_INIT.
 DR SWISS-2DPAGE; P26428; COLI.
 DR Ecogene; E611383; elbB.
 DR InterPro; IPR002818; ThiJ.
 DR Pfam; PF01965; Df-1_Pfpl; 1.
 KM Isoprene biosynthesis; Complete proteome.
 FT CONFLICT 111 111
 FT C->S (IN REF. 1).
 FT CONFLICT 195 217
 FT PACFWLNE (IN REF. 1).
 FT KONTAKASGIDKLSRVLVLAIE -> RRTLOKRAALISM
 SQ SEQUENCE 217 AA; 22981 MM; 48A7957C29384DAC CRC64;

Query Match 1.6%; Score 7; DB 1; Length 217;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 270 KLVSRVL 276
 |||||
 DB 207 KLVSRVL 213

RESULT 15
 R103 ORYSA STANDARD; PRT; 219 AA.
 ID R103 ORYSA
 AC Q40649; Q40717;
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE 60S ribosomal protein L10-3 (OM/R22).
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=CV. INDICA GUANG-LU-AI NO.4;
 RA Zong H.;
 RL Thesis (1996), Fudan University, China.
 [2]
 RN SEQUENCE OF 77-219 FROM N.A.
 RC STRAIN=CV. Indica-IR36; TISSUE=Seed;
 RA Kim J.K.;
 RT Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE L10E FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC -----
DR EMBL; U55048; AAA98698.1; -.
DR EMBL; X64621; CAA45905.1; -.
DR PIR; S19224; S19224.
DR Gramene; Q40649; -.
DR InterPro; IPR001197; Ribosomal_L10E.
DR Pfam; PF00826; Ribosomal_L10E; 1.
DR TIGRPFAMs; TIGR00279; L10E; 1.
DR PROSITE; PS01257; RIBOSOMAL_L10E; 1.
KM Ribosomal protein.
PT CONFLICT 77 TKS -> EFP (IN REF. 2).
SQ SEQUENCE 219 AA; 24672 MW; CE57477068725DBF CRC64;

```

```

Query Match 1.6%; Score 7; DB 1; Length 219;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 7 AFGKAPL 13
DB 210 AFGKAPL 216

```

```

RESULT 16
ID RL10_MAIZE STANDARD; PRT; 220 AA.
AC P45633;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 60S ribosomal protein L10 (QM protein homolog).
GN RPL10.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OC NCBI_TaxID=4577;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. B73;
RX MEDLINE=94362669; PubMed=8081358;
RA Neill J.D.;
RT "Extreme evolutionary conservation of QM, a novel c-Jun associated
transcription factor."
RL Hum. Mol. Genet. 3:723-728 (1994).
CC -1- SIMILARITY: BELONGS TO THE L10E FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U06108; AAA17419.1; -.
DR PIR; T02068; T02068.
DR MaizeDB; 77933; -.
DR InterPro; IPR001197; Ribosomal_L10E.
DR Pfam; PF00826; Ribosomal_L10E; 1.
DR TIGRPFAMs; TIGR00279; L10E; 1.
DR PROSITE; PS01257; RIBOSOMAL_L10E; 1.
KM Ribosomal protein.
SQ SEQUENCE 220 AA; 24919 MW; 5B338061E25893B8 CRC64;

```

```

Query Match 1.6%; Score 7; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 7 AFGKAPL 13
DB 210 AFGKAPL 216

```

```

RESULT 17
ID CDSA_RICPR STANDARD; PRT; 228 AA.
AC Q9ZDA8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphatidate cytidyltransferase (EC 2.7.7.41) (CDP-diacylglycerol
synthetase) (CDP-diacylglycerol pyrophosphorylase) (CDP-diacylglycerol
synthase) (CDP-diacylglycerol synthetase) (CDP-DAG
synthase) (CDP-DG synthetase).
GN CDSA OR RP424.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.B.; Zomorodipour A.; Andersson J.O.;
RA Scharitz-Ponten T.; Almark U.C.M.; Podowski R.M.; Naeslund A.K.;
RA Eriksson A.-S.; Winkler H.H.; Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
mitochondria."
RL Nature 396:133-140 (1998).
CC -1- CATALYTIC ACTIVITY: CTP + phosphatidate = diphosphate + CDP-
diacylglycerol.
CC -1- PATHWAY: Phospholipid biosynthesis.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (by similarity).
CC -1- SIMILARITY: BELONGS TO THE CDS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ235271; CAA14881.1; -.
DR PIR; G71700; G71700.
DR InterPro; IPR000374; PCtransf.
DR Pfam; PF01148; CTP_transf_1; 1.
DR PROSITE; PS01315; CDS; 1.
KM Transferrase; Nucleotidyltransferase; Phospholipid biosynthesis;
KW Transmembrane; Complete proteome.
FT TRANSMEM 31 51 POTENTIAL.
FT TRANSMEM 65 85 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 131 151 POTENTIAL.
FT TRANSMEM 165 185 POTENTIAL.
FT TRANSMEM 206 226 POTENTIAL.
SQ SEQUENCE 228 AA; 25814 MW; 4BDE6A6BA19174 CRC64;

```

```

Query Match 1.6%; Score 7; DB 1; Length 228;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 150 SAGLVSV 156
DB 140 SAGLVSV 146

```

```

RESULT 18
ID KKA3_ENTFA STANDARD; PRT; 264 AA.
AC P00554;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aminoglycoside 3'-phosphotransferase (EC 2.7.1.95) (Kanamycin kinase,
type III) (Neomycin-kanamycin phosphotransferase, type III)

```

DE (APH(3')III).
 GN APHA.
 OS Enterococcus faecalis (Streptococcus faecalis), and
 OS Staphylococcus aureus.
 OG Plasmid pJH1.
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
 OX NCBI_TaxID=1351, 1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=E.faecalis;
 RX MEDLINE=84029883; PubMed=6313476;
 RA Trien-Cout P., Courvalin P.,
 RT "Nucleotide sequence of the Streptococcus faecalis plasmid gene
 RL encoding the 3'5'-aminoglycoside phosphotransferase type III.",
 RN Gene 23:331-341(1983).
 [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.aureus;
 RX MEDLINE=88174293; PubMed=6100986;
 RA Gray G.S., Fitch W.M.,
 RT "Evolution of antibiotic resistance genes: the DNA sequence of a
 RL kanamycin resistance gene from Staphylococcus aureus.",
 RN Mol. Biol. Evol. 1:57-66(1983).
 CC -1- FUNCTION: RESISTANCE TO KANAMYCIN AND STRUCTURALLY RELATED
 CC AMINOGLYCOSIDES, INCLUDING AMIKACIN.
 CC -1- CATALYTIC ACTIVITY: ATP + kanamycin = ADP + kanamycin 3'-
 CC phosphate.
 CC -1- SIMILARITY: TO OTHER AMINOGLYCOSIDE PHOSPHOTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; V01547; CAA24789.1; -.
 DR EMBL; M36771; AAA26596.1; ALT_SEQ.
 DR EMBL; X93398; CAA67773.1; -.
 DR PIR; A00665; PKSOJF.
 DR PIR; A26220; PKSAF.
 DR PDB; 1J7I; 08-AUG-01.
 DR PDB; 1J7L; 08-AUG-01.
 DR PDB; 1J7U; 08-AUG-01.
 DR PDB; 1L8T; 19-JUN-02.
 DR PDB; 1L8U; 19-JUN-02.
 DR InterPro; IPR002575; APH.
 DR Pfam; PF01636; APH; 1.
 KW Antibiotic resistance; Transferase; Kinase; ATP-binding; Plasmid;
 KM 3D-structure.
 FT ACT SITE 190 190 BY SIMILARITY.
 FT VARIANTS 35 35 MISSING (IN S.AUREUS).
 FT SEQUENCE 264 AA; 30974 MW; 7EEA5851D3380C5A CRC64;
 SQ
 Query Match 1.6%; Score 7; DB 1; Length 264;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 233 ELKULIE 239
 DB 9 ELKULIE 15
 RESULT 19
 ID RFAP_SALTY STANDARD; PRT; 265 AA.
 AC Q06995; O68268;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Lipopolysaccharide core biosynthesis protein rfap.
 GN RFAP OR WAAP OR STM3721.

OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2;
 RX MEDLINE=98204873; PubMed=9535865;
 RA Heinrichs D.E., Monteiro M.A., Perry M.B., Whitfield C.,
 RT "The assembly system for the lipopolysaccharide R2 core-type of
 RT Escherichia coli is a hybrid of those found in Escherichia coli K-12
 RT and Salmonella enterica. Structure and function of the R2 waak and
 RL Waal homologs.",
 RN J. Biol. Chem. 273:8849-8859(1998).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E.,
 RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.,
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.",
 RN Nature 413:852-856(2001).
 [3]
 RP SEQUENCE OF 177-265 FROM N.A.
 RC STRAIN=LT2;
 RX MEDLINE=93186722; PubMed=8444813;
 RA Kleina J.D., Pradel E., Schaitman C.A.,
 RT "The rfa gene, which is involved in production of a rough form of
 RT lipopolysaccharide core in Escherichia coli K-12, is not present in
 RL the rfa cluster of Salmonella typhimurium LT2.",
 CC J. Bacteriol. 175:1524-1527(1993).
 CC -1- FUNCTION: INVOLVED IN ATTACHMENT OF PHOSPHATE-CONTAINING
 CC SUBSTITUENTS TO THE INNER CORE.
 CC -1- PATHWAY: Lipopolysaccharide core biosynthesis.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF026386; AAC16414.1; -.
 DR EMBL; AE008873; AAL22580.1; -.
 DR EMBL; S56361; AAB25550.1; -.
 DR PIR; A47074; A47074.
 DR ScyGene; SG10341; rfap.
 DR Pfam; PF05260; Waap; 1.
 KW Lipopolysaccharide biosynthesis; Complete proteome.
 FT CONFLICT 224 224 F -> Y (IN REF. 3).
 FT SEQUENCE 265 AA; 30996 MW; 903BFP9C319667F30 CRC64;
 SQ
 Query Match 1.6%; Score 7; DB 1; Length 265;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 42 KGTSLKE 48
 DB 53 KGTSLKE 59
 RESULT 20
 ID PTRC_KLEPN STANDARD; PRT; 266 AA.
 AC P37082;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

```

DE PTS system, sorbose-specific IIC component (EIIIC-SOR) (Sorbose-
DE permease IIC component) (Phosphotransferase enzyme II, C component).
GN SORA.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1033-5P14 / KAY2026;
RX MEDLINE=95035038; PubMed=7947966;
RA Wehmeler U.R., Lengeler J.W.;
RA "Sequence of the sor-operon for L-sorbose utilization from Klebsiella
RT pneumoniae KAY2026."
RL Biochim. Biophys. Acta 1208:348-351(1994).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -1- SIMILARITY: Contains 1 PTS EIIIC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X66059; CA46859.1; -.
DR PIR, S50189; S50189.
DR InterPro; IPR004700; PTSIIC_sorbose.
DR Pfam; PF03609; EII-Sor; 1.
DR TIGRPFAMs; TIGR00882; EII-Sor; 1.
KW Phosphotransferase system; Sugar transport; Transmembrane;
KW Inner membrane.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 33 53 POTENTIAL.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 100 120 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 183 203 POTENTIAL.
FT TRANSMEM 219 239 POTENTIAL.
SQ SEQUENCE 266 AA; 27872 MW; 504FD32A4F268FBC CRC64;

Query Match 1.6%; Score 7; DB 1; Length 266;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -1- FUNCTION: Seems to use temperature-induced exposure of structured
CC hydrophobic domains to capture early unfolding protein
CC intermediates and rapidly release them in an active form once
CC stress has abated (by similarity).
CC -1- SIMILARITY: Belongs to the hcpA family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AE004543; MAG04524.1; -.
DR PIR, C83504; C83504.
DR HAMAP; MF_01046; -.
DR InterPro; IPR002818; ThiJ.
DR Pfam; PF01965; Df-1_Pfpi; 1.
KW Chapterone; Complete proteome.
SQ SEQUENCE 291 AA; 31898 MW; B55BCD412BE4AA0A CRC64;

Query Match 1.6%; Score 7; DB 1; Length 291;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 423 ECKPEDY 429
200 ECKPEDY 206

RESULT 22
NADA_FUSNN STANDARD; PRT; 298 AA.
AC Q8RCG9;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Quinolinate synthetase A.
GN NADA OR FN0008.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 25586;
RX MEDLINE=N-2186394; PubMed=11889109;
RA Kapralin V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Greenkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Gotsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fongsein M., Kyriades N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
CC -1- FUNCTION: Catalyzes the condensation of iminoacetate with
CC dihydroxyacetone phosphate to form quinolinate (By similarity).
CC -1- PATHWAY: NAD biosynthesis; aspartate to NADM; second step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE QUINOLINATE SYNTHETASE A FAMILY.

```

```

Db 12 FLAGGYL 18
214 FLAGGYL 220

RESULT 21
HCHA_PSEAE STANDARD; PRT; 291 AA.
AC Q914K0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Chapterone protein hcha (Hsp31).
GN HCHA OR PA1135.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]

```


CC SUBFAMILY 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

DR EMBL: A010516; AAL94221.1; -.

DR InterPro: IPR003473; NADL.

DR Pfam: PF02445; NADA; 1.

DR TIGR: TIGR00550; nada; 1.

KW Pyridine nucleotide biosynthesis; Complete proteome.

SO SEQUENCE 298 AA; 33813 MW; B92353C2ADE3801 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 298;

Best Local Similarity 100.0%; Pred. No. 40;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 EMBEYD 312

DB 94 EMBEYD 100

RESULT 23

FLAA BORBU STANDARD; PRT; 341 AA.

AC P70856; O51612; Q44876;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE Flagellar filament outer layer protein precursor (sheath protein).

GN FLAA OR BB0668.

OS Borrelia burgdorferi (Lyme disease spirochete).

OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.

OX NCBI_TaxID=139;

RA NCB1_TaxID=139;

RP SEQUENCE FROM N.A.

RC STRAIN=212;

RX MEDLINE=97144545; PubMed=8990312;

RT Ge Y., Charon N.;

RT "An unexpected flaa homolog is present and expressed in Borrelia

RT burgdorferi.";

RT J. Bacteriol. 179:552-556(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 35210 / B31;

RX MEDLINE=98065943; PubMed=9403685;

RA Fraser C.M., Castjens S., Huang W.M., Sutton G.G., Clayton R.A.,

RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,

RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,

RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,

RA Ueberback T., Wathey L., McDonald L., Attiach P., Bowman C.,

RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,

RA Smith H.O., Venter J.C.;

RA "Genomic sequence of a Lyme disease spirochete, Borrelia

RA burgdorferi.";

RT Nature 390:580-586(1997).

RN [3]

RP SEQUENCE OF 276-341 FROM N.A.

RC STRAIN=212;

RX MEDLINE=98438936; PubMed=9765799;

RA Old I.G., Trueba G.A., Saint-Girons I., Johnson R.C.;

RA "A chea chem operon in Borrelia burgdorferi, the agent of Lyme

RA disease.";

RT Res. Microbiol. 148:191-200(1997).

CC -1- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.

CC -1- SUBUNIT: THE FLAGELLUM CONSISTS OF AN OUTER LAYER COMPOSED OF

CC REPEATING UNITS OF FLAA AROUND A CORE THAT CONTAINS SEVERAL

CC ANTIGENICALLY RELATED POLYPEPTIDES.

CC -1- SUBCELLULAR LOCATION: Periplasmic flagellum.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

DR EMBL: U62900; AAC44770.1; -.

DR EMBL: X901166; AAC67025.1; ALT_INIT.

DR EMBL: X91907; CAA63001.1; -.

DR TIGR: BB0668; -.

DR InterPro: IPR006714; FLAA.

DR Pfam: PF04620; FLAA; 1.

KW Flagella; Periplasmic; Signal; Complete proteome.

FT SIGNAL 1 19

FT CHAIN 20 341

FT CONFLICT 254 255

FT CONFLICT 317 318

SO SEQUENCE 341 AA; 38441 MW; 60B1475EBADP3451 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 341;

Best Local Similarity 100.0%; Pred. No. 45;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 EXPRESS 287

DB 332 EXPRESS 338

RESULT 24

UN97_CABEL STANDARD; PRT; 348 AA.

ID UN97_CABEL

AC P50464;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE LIM protein unc-97.

GN UNC-97 OR FLAD12.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=99102422; PubMed=9885243;

RA Hobert O., Moerman D.G., Clark K.A., Beckerle M.C., Ruvkun G.;

RA "A conserved LIM protein that affects muscular adhesion junction

RA integrity and mechanosensory function in Caenorhabditis elegans.";

RT J. Cell Biol. 144:45-57(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Mink P.;

RA Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: PROBABLY FUNCTION IN ADHERENS JUNCTION. UNC-97 AFFECTS

CC THE STRUCTURAL INTEGRITY OF THE INTEGRIN CONTAINING MUSCLE

CC ADHERENS JUNCTIONS AND CONTRIBUTES TO THE MECHANOSENSORY FUNCTIONS

CC OF TOUCH NEURONS.

CC -1- SUBCELLULAR LOCATION: ADHERENS JUNCTIONS AND ALSO NUCLEAR.

CC -1- TISSUE SPECIFICITY: RESTRICTED TO TISSUE TYPES THAT ATTACH TO THE

CC HYPODERMIS, SPECIFICALLY BODY WALL MUSCLES, VULVAL MUSCLES, AND

CC MECHANOSENSORY NEURONS.

CC -1- SIMILARITY: Contains 5 LIM zinc-binding domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AF035583; AAD09435.1; -.

DR EMBL: U01021; AAA82334.1; -.

DR PIR: T16076; T16076.

DR HSSP: P48059; 1G47.

DR WormPep: F14D12.2; CE04392.

DR InterPro: IPR001781; LIM.

DR Pfam: PF00412; LIM; 5.

DR ProDom: PD000094; LIM; 5.

DR SMART: SM00132; LIM; 5.

DR PROSITE: PS00478; LIM DOMAIN 1; 4.

DR PROSITE: PS50023; LIM DOMAIN 2; 5.

KM LIM domain; Metal-binding; Zinc; Repeat.

FT DOMAIN 21 73 LIM 1.

FT DOMAIN 82 132 LIM 2.

FT DOMAIN 146 196 LIM 3.

FT DOMAIN 205 255 LIM 4.

FT DOMAIN 264 315 LIM 5.

SEQUENCE 348 AA; 40308 MW; 0CDD2C39628CEDA CRC64;

Query Match

Best Local Similarity 1.6%; Score 7; DB 1; Length 348;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 230 PFTELKK 236

DB 319 PFTELKK 325

RESULT 25

SOTB_HAEIN STANDARD; PRT; 396 AA.

AC P44535;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Probable sugar efflux transporter.

GN SOTB OR H10135.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Haemophilus.

OX NCBI_TaxID=727;

DB [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RD / KW20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom B., Cotton M.D., Utechtack T.R., Hanna M.C., Nguyen D.T., Saudke D.M., Brandon R.C., Rabe L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M., RAE Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;

RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";

RL Science 269:496-512 (1995).

CC -1- FUNCTION: Involved in the efflux of sugars. The physiological role may be the reduction of the intracellular concentration of toxic sugars or sugar metabolites (By similarity).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (Probable).

CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN AS THE DRUG RESISTANCE TRANSLUCASE FAMILY). SOTB (TC 2.A.1.2)

CC SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U32699; AAC21806.1; -.

DR PIR: E64143; E64143.

DR TIGR: H10135; -.

DR HAMAP: MF_00517; -; 1.

DR InterPro: IPR007114; MFS.

DR InterPro: IPR005828; Sub transporter.

DR Pfam: PF00083; Sugar_tr; 1.

KW Transport; Sugar transporter; Transmembrane; Inner membrane; Complete proteome.

FT TRANSMEM 15 35 POTENTIAL.

FT TRANSMEM 51 71 POTENTIAL.

FT TRANSMEM 84 104 POTENTIAL.

FT TRANSMEM 109 129 POTENTIAL.

FT TRANSMEM 137 157 POTENTIAL.

FT TRANSMEM 168 188 POTENTIAL.

FT TRANSMEM 209 229 POTENTIAL.

FT TRANSMEM 245 265 POTENTIAL.

FT TRANSMEM 273 293 POTENTIAL.

FT TRANSMEM 297 317 POTENTIAL.

FT TRANSMEM 333 353 POTENTIAL.

FT TRANSMEM 365 385 POTENTIAL.

SEQUENCE 396 AA; 43360 MW; B7994F2F9E9339C8 CRC64;

Query Match

Best Local Similarity 1.6%; Score 7; DB 1; Length 396;

Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 197 DVATAY 203

DB 332 DVATAY 338

RESULT 26

ASSY_XYLFA STANDARD; PRT; 401 AA.

AC O9PM9;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Argininosuccinate synthase (EC 6.3.4.5) (Citruiline--aspartate ligase).

GN ARG3 OR XP0999.

OS Xylella fastidiosa.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

OC Xanthomonadaceae; Xylella.

OX NCBI_TaxID=2371;

DB [1]

RP SEQUENCE FROM N.A.

RC STRAIN=9A5C.

RX MEDLINE=20565717; PubMed=10910347;

RA Simpson A.J.G., Reinach F.C., Artida P., Abreu F.A., Acencio M., Alvarez A.R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S., Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Bioness M.R.S., Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H., Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M., Coutinho L.L., Cristofani M., Dias-Neto B., Docena C., El-Dorri H., Fachincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.R., Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R., Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A., Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P., Krieger J.E., Kuramae E.B., Laigret F., Lambais M.R., Leite L.C.C., Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A., Marques M.V., Martins A.M.B.N., Madeira A.M.F., Martino C.L., Marques C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B., Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S., Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A., de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A., Peixoto B.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,

```

RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.B. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Trai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-citrulline + L-aspartate = AMP +
CC diphosphate + L-argininosuccinate.
CC -1- PATHWAY: Arginine biosynthesis; seventh step.
CC -1- SUBUNIT: Homotrimer (by similarity).
CC -1- SIMILARITY: Belongs to the argininosuccinate synthase family.
Subfamily 1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AE003937; AAF83809.1; -.
CC PIR; G82737; G82737.
CC HAMAP; MF_00005; -.
CC InterPro; IPR001518; Arginosuc_synth.
CC Pfam; PF00764; Arginosuc_synth; 1.
CC ProDom; PD003544; Arginosuc_synth; 1.
CC TIGRfams; TIGR00032; arg9; 1.
CC PROSITE; PS00564; ARGININOSUCCIN SYN 1; 1.
CC PROSITE; PS00565; ARGININOSUCCIN SYN 2; FALSE NEG.
CC Arginine biosynthesis; ligase; ATP-binding; Complete proteome.
KW SEQUENCE 401 AA; 44017 MW; CESFOD74B356466D CRC64;
SQ
Query Match 1.6%; Score 7; DB 1; Length 401;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 328 PLTVAIK 334
DB 216 PLTVAIK 222
ID ASXY_XYLFT STANDARD; PRT; 401 AA.
AC P59606;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Argininosuccinate synthase (EC 6.3.4.5) [Citrulline-aspartate
DE ligase].
DE ARG OR PD0291.
OS Xylella fastidiosa (strain Temecula / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=2424331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Pietro M.I.T., da Silva F.R.,
RA Goldman W.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carrier H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Pietro E.S., Haraoka R., Kuramae E.B.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Camavan F.S., Celestino A.V.,
RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leon S.G., Oliveira A.R., Rosa V.B. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsubako M., Yama G.M., Zaros L.G.,

```

```

RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa."
RL J. Bacteriol. 185:1018-1026(2003).
CC -1- CATALYTIC ACTIVITY: ATP + L-citrulline + L-aspartate = AMP +
CC diphosphate + L-argininosuccinate.
CC -1- PATHWAY: Arginine biosynthesis; seventh step.
CC -1- SUBUNIT: Homotrimer (by similarity).
CC -1- SIMILARITY: Belongs to the argininosuccinate synthase family.
Subfamily 1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AE012554; AAO28176.1; -.
CC HAMAP; MF_00005; -.
CC ProDom; PD003544; Arginosuc_synth; 1.
CC PROSITE; PS00564; ARGININOSUCCIN SYN 1; 1.
CC PROSITE; PS00565; ARGININOSUCCIN SYN 2; FALSE NEG.
CC Arginine biosynthesis; ligase; ATP-binding; Complete proteome.
KW SEQUENCE 401 AA; 43970 MW; A37AB801A22F4212 CRC64;
SQ
Query Match 1.6%; Score 7; DB 1; Length 401;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 328 PLTVAIK 334
DB 216 PLTVAIK 222
ID KIME_SCHPO STANDARD; PRT; 404 AA.
AC Q09780;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative mevalonate kinase (EC 2.7.1.36) (MK).
DE SPAC1366.11C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsis K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voiclaert G., Aert R., Robben J., Grympos B.,
RA Welljens I., Vanstreels E., Rieger M., Scheifer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinrich H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

```

RA Goffeau A., Cadiou E., Dreano S., Gloux S., Ielaure V., Mottier S.,
 RA Gallibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Luczak M., Rochet S., Galliard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 RT "The genome sequence of *Schizosaccharomyces pombe*."
 RN Nature 415:871-880(2002).
 RN [12]
 RP SEQUENCE FROM N.A.
 RX STRAIN=PR745;
 RX MEDLINE=98162722; PubMed=9501991;
 RA Yoshoka S., Kato K., Nakai K., Okayama H., Nojima H.;
 RT "Identification of open reading frames in *Schizosaccharomyces pombe*
 RT CDNA8."
 RL DNA Res. 4:363-369(1997).
 CC -1- CATALYTIC ACTIVITY: ATP + (R)-nevalonate = ADP + (R)-5-
 CC phosphomevalonate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO THE GMP KINASE FAMILY. MEVALONATE KINASE
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z54308; CAA91104.1; -;
 DR EMBL; AB000541; BAA25169.1; -;
 DR PIR; S62440; S62440.
 DR GeneDB_Spombe; SPAC13G6.11c; -;
 DR InterPro; IPR001174; Galkinase.
 DR InterPro; IPR006204; GMP_kinase.
 DR InterPro; IPR006203; GMP_kinase ATP.
 DR InterPro; IPR006205; Mv_gal_kin.
 DR InterPro; IPR006206; Mv_gal_kinase.
 DR Pfam; PF00288; GMP_kinase8.1.
 DR PRINTS; PR00960; LM6PROTEIN.
 DR PRINTS; PR00959; MEVGALKINASE.
 DR TIGRFAMs; TIGR00549; mevalon_kin.1.
 DR PROSITE; PS00627; GMP_KINASES ATP; 1.
 KW Hypothetical protein; Transferase; Kinase; Sterol biosynthesis;
 KW ATP-binding.
 FT NE_BIND 133 143 ATP (BY SIMILARITY).
 SQ SEQUENCE 404 AA; 43406 MW; AEBBD61646247710 CRC64;
 QY 156 VVATSL 162
 DB 147 VVATSL 153
 Query Match 1.6%; Score 7; DB 1; Length 404;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RP SEQUENCE FROM N.A.
 RX MEDLINE=92093615; PubMed=1684419;
 RA Faiella A., D'Esposito M., Rambaldi M., Acampora D.,
 RA Ballofiore S., Stornaiuolo A., Mallamaci M., Migliaccio E.,
 RA Gullisano M.;
 RT "Isolation and mapping of EVX1, a human homeobox gene homologous to
 RT even-skipped, localized at the 5' end of HOX1 locus on chromosome
 RT 7."
 RL Nucleic Acids Res. 19:6541-6545(1991).
 RN [12]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96070899; PubMed=7499236;
 RA Briata P., van de Werken R., Altoidi I., Ilengo C., di Bias E.,
 RA Boncinelli E., Corte G.;
 RT "Transcriptional repression by the human homeobox protein EVX1 in
 RT transfected mammalian cells."
 RL J. Biol. Chem. 270:27695-27701(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Bradshaw H., Hinds K., Keppler D.;
 RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE SPECIFICATION OF NEURONAL CELL
 CC TYPES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE EVEN-SKIPPED HOMEBOX FAMILY.
 CC PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X60655; CAA43062.1; -;
 DR EMBL; U68782; AAB07598.1; -;
 DR EMBL; U68781; AAB07598.1; JOINED.
 DR EMBL; AC004080; -; NOT_ANNOTATED_CDS.
 DR PIR; S22586; S22586.
 DR HSSP; P14653; 1B72.
 DR TRANSFAC; T02021; -;
 DR GeneW; HGNC:3506; EVX1.
 DR MIM; 142996; -;
 DR GO; GO:0003700; P:transcription factor activity; TAS.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox.1.
 DR PRINTS; PR00024; HOMEBOX.
 DR ProDom; PD000010; Homeobox.1.
 DR SMART; SMO0389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX 1; 1.
 DR PROSITE; PS00071; HOMEBOX 2; 1.
 KW DNA-binding; Developmental protein; Homeobox; Nuclear protein.
 FT DOMAIN 114 124 ASP/GLU-RICH (ACIDIC).
 FT DNA_BIND 183 242 HOMEBOX.
 FT DOMAIN 289 297 ALA-RICH.
 FT DOMAIN 361 368 ALA-RICH.
 SQ SEQUENCE 407 AA; 42439 MW; 75D696ECA1431D5A CRC64;
 QY 264 GSETPK 270
 DB 155 GSETPK 161
 Query Match 1.6%; Score 7; DB 1; Length 407;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 29
 EVX1_HUMAN
 ID EVX1_HUMAN STANDARD; PRT; 407 AA.
 AC P49640;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Homeobox even-skipped homolog protein 1 (EVX-1).
 GN EVX1.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_taxid=9606;
 RN [1]

RESULT 30
 PSD4_SCHEMA
 ID PSD4_SCHEMA STANDARD; PRT; 420 AA.
 AC O17453;
 DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 26S proteasome non-ATPase regulatory subunit 4 (26S proteasome
 DE regulatory subunit 55A).
 OS Schistosoma mansoni (Blood fluke).
 CC Schistosoma mansoni (Blood fluke).
 CC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeididae;
 CC Schistosomatidae; Schistosomatidae; Schistosoma.
 NCBI_TaxID=6183;
 OK NCBI_TaxID=6183;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Puerto Rican;
 RA Harrop R., Wilson R.A.;
 RL Submitted (Oct-1997) to the EMBL/GenBank/DDBJ databases.
 CC -1- FUNCTION: BINDS AND PRESUMABLY SELECTS UBQUITIN-CONJUGATES FOR
 CC DESTRUCTION (BY SIMILARITY).
 CC -1- SUBUNIT: THE 26S PROTEASOME IS COMPOSED OF A CORE PROTEASE, KNOWN
 CC AS THE 20S PROTEASOME, CAPPED AT ONE OR BOTH ENDS BY THE 19S
 CC REGULATORY COMPLEX (RC). THE RC IS COMPOSED OF AT LEAST 18
 CC DIFFERENT SUBUNITS IN TWO SUBCOMPLEXES, THE BASE AND THE LID,
 CC WHICH FORM THE PORTIONS PROXIMAL AND DISTAL TO THE 20S PROTEOLYTIC
 CC CORE, RESPECTIVELY (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE PROTEASOME SUBUNIT 55A FAMILY.
 CC -1- SIMILARITY: Contains 3 ubiquitin-interacting motif (UIM) repeats.
 CC -1- SIMILARITY: Contains 1 WFMA domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF030960; AAB86561.1; ALT_INIT.
 DR InterPro: IPR003903; UIM.
 DR InterPro: IPR002035; WFMA.
 DR Pfam: PF02809; UIM; 3.
 DR SMART: SM00726; UIM; 3.
 DR SMART: SM00327; VMA; 1.
 DR PROSITE: PS50330; UIM; 1.
 DR PROSITE: PS50234; WFMA; FALSE_NEG.
 KM Proteasome; Repeat.
 FT DOMAIN 1 174 WFMA.
 FT DOMAIN 210 229 UIM 1.
 FT DOMAIN 255 274 UIM 2.
 FT DOMAIN 288 307 UIM 3.
 SQ SEQUENCE 420 AA; 45728 MW; 441F8D0303B7ACD1 CRC64;
 Query Match 1.6%; Score 7; DB 1; Length 420;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 155 SVVATSL 161
 DB 242 SVVATSL 248
 RESULT 31
 YG84_METTH STANDARD; PRT; 431 AA.
 AC 027719;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MTH1684.
 GN MTH1684.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 CC NCBI_TaxID=187420;
 CC [1]
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Delta H;

RX MEDLINE=98037514; PubMed=93771463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee R.-M., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakey D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jivani N., Caruso A., Bush D., Sater H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pletrovski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT delta: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155 (1997).
 CC -1- SIMILARITY: STRONG, TO M.JANNA-SCHIT MJ1681.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000926; AAB86156.1;
 DR PIR: B69092; B69092.
 DR HSSP: P0198; 2FDN.
 DR InterPro: IPR001450; 4Fe4S_ferredoxin.
 DR InterPro: IPR002708; DUF39.
 DR Pfam: PF01837; DUF39; 1.
 DR Pfam: PF00037; Fe4S; 2.
 DR ProDom: PD011569; DUF39; 1.
 DR PROSITE: PS00198; 4Fe4S_FERREDOXIN; UNKNOWN 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 431 AA; 46950 MW; 802FA39757194B85 CRC64;
 Query Match 1.6%; Score 7; DB 1; Length 431;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 233 ELKRLIE 239
 DB 418 ELKRLIE 424
 RESULT 32
 AMPA_HELPJ STANDARD; PRT; 496 AA.
 ID AMPA_HELPJ
 AC 092181;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase) (LAP)
 DE (Leuyl aminopeptidase).
 GN PEPA OR UHP0517.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteriaceae; Helicobacter.
 CC NCBI_TaxID=85963;
 CC [1]
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Molir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori.";
 RL Nature 397:176-180 (1999).
 CC -1- FUNCTION: PRESUMABLY INVOLVED IN THE PROCESSING AND REGULAR
 CC TURNOVER OF INTRACELLULAR PROTEINS. CATALYZES THE REMOVAL OF
 CC UNSUBSTITUTED N-TERMINAL AMINO ACIDS FROM VARIOUS PEPTIDES (BY
 CC SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-
 CC Xbb-, in which Xaa is preferably Leu, but may be other amino acids

including Pro although not Arg or Lys, and Xbb may be Pro.
 CC -1- COFACTOR: Binds 2 manganese ions per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M17.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE001485; AAD06098.1; -.
 CC PIR; P00727; P71920.
 CC HSSP; P00727; 1LAP.
 CC HAMAP; MF_00181; -; 1.
 CC InterPro; IPR000819; Peptidase_M17.
 CC Pfam; PF00883; Peptidase_M17; 1.
 CC PRINTS; PR00481; LAMNOPTDASE.
 CC PROSITE; PS00631; CYTOSOL_AP; 1.
 CC KM Hydroxylase; Aminopeptidase; Manganese; Complete proteome.
 CC FT METAL 258 258 MANGANESE 2 (BY SIMILARITY).
 CC FT METAL 263 263 MANGANESE 1 AND 2 (BY SIMILARITY).
 CC FT METAL 281 281 MANGANESE 2 (BY SIMILARITY).
 CC FT METAL 340 340 MANGANESE 1 (BY SIMILARITY).
 CC FT METAL 342 342 MANGANESE 1 AND 2 (BY SIMILARITY).
 CC FT ACT SITE 270 270 POTENTIAL.
 CC FT ACT SITE 344 344 POTENTIAL.
 CC SQ SEQUENCE 496 AA; 54611 MW; 768CCE1520A0EE CRC64;
 CC
 CC Query Match 1.6%; Score 7; DB 1; Length 496;
 CC Best Local Similarity 100.0%; Pred. No. 62;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 234 LKLTIES 240
 CC DB 414 LKLTIES 420
 CC
 CC RESULT 33
 CC AMPA_HELPY STANDARD; PRT; 496 AA.
 CC ID 025294;
 CC DT 15-JUL-1998 (Rel. 36, Created)
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase) (LAP)
 CC DB (leucyl aminopeptidase).
 CC GN PEPA OR HP0570.
 CC OS Helicobacter pylori (Campylobacter pylori).
 CC OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 CC OC Helicobacteriaceae; Helicobacter.
 CC OX NCBI_Taxid=210;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=26695 / ATCC 700392;
 CC RX MEDLINE=97394467; PubMed=9252185;
 CC RA Tomb J.-F., White O., Keriavage A.R., Clayton R.A., Sutton G.G.,
 CC RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 CC RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 CC RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 CC RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 CC RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 CC RA Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
 CC RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 CC RA Venter J.C.;
 CC RT "The complete genome sequence of the gastric pathogen Helicobacter
 CC RT pylori.";
 CC RL Nature 388:539-547(1997).
 CC -1- FUNCTION: PRESUMABLY INVOLVED IN THE PROCESSING AND REGULAR
 CC CC TURNOVER OF INTRACELLULAR PROTEINS. CATALYZES THE REMOVAL OF
 CC CC UNSUBSTITUTED N-TERMINAL AMINO ACIDS FROM VARIOUS PEPTIDES (BY

SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-
 CC Xbb-, in which Xaa is preferably Leu, but may be other amino acids
 CC including Pro although not Arg or Lys, and Xbb may be Pro.
 CC -1- COFACTOR: Binds 2 manganese ions per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M17.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000571; AAD07638.1; -.
 CC PIR; B64591; B64591.
 CC HSSP; P00727; 1BPV.
 CC DR MEROPS; M17.UPM; -.
 CC DR TIGR; HP0570; -.
 CC HAMAP; MF_00181; -; 1.
 CC InterPro; IPR000819; Peptidase_M17.
 CC DR Pfam; PF00883; Peptidase_M17; 1.
 CC PRINTS; PR00481; LAMNOPTDASE.
 CC PROSITE; PS00631; CYTOSOL_AP; 1.
 CC KM Hydroxylase; Aminopeptidase; Manganese; Complete proteome.
 CC FT METAL 258 258 MANGANESE 2 (BY SIMILARITY).
 CC FT METAL 263 263 MANGANESE 1 AND 2 (BY SIMILARITY).
 CC FT METAL 281 281 MANGANESE 2 (BY SIMILARITY).
 CC FT METAL 340 340 MANGANESE 1 (BY SIMILARITY).
 CC FT METAL 342 342 MANGANESE 1 AND 2 (BY SIMILARITY).
 CC FT ACT SITE 270 270 POTENTIAL.
 CC FT ACT SITE 344 344 POTENTIAL.
 CC SQ SEQUENCE 496 AA; 54433 MW; 565A0BCA825FC7 CRC64;
 CC
 CC Query Match 1.6%; Score 7; DB 1; Length 496;
 CC Best Local Similarity 100.0%; Pred. No. 62;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 234 LKLTIES 240
 CC DB 414 LKLTIES 420
 CC
 CC RESULT 34
 CC CPT7_ORYLA STANDARD; PRT; 517 AA.
 CC ID CPT7_ORYLA
 CC AC P70085; P70086;
 CC DT 15-DEC-1998 (Rel. 37, Created)
 CC DT 15-DEC-1998 (Rel. 37, Last sequence update)
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
 CC DE Cytochrome P450 17A1 (EC 1.14.99.9) (CYP17A1) (P450-C17) (Steroid 17-
 CC DE alpha-hydroxylase/17,20 lyase).
 CC GN CYP17A1 OR CYP17.
 CC OS Oryzias latipes (Medaka fish) (Japanese ricefish).
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Neuteleostei;
 CC OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neuteleostei;
 CC OC Acanthomorpha; Acanthopterygii; Percomorph; Atherinomorpha;
 CC OC Belontiiformes; Atherinichthyidae; Oryziinae; Oryzias.
 CC OX NCBI_Taxid=8090;
 CC RN [1]
 CC RP SEQUENCE FROM N.A. (ISOFORMS C17L AND C17S).
 CC RC STRAIN=Orange-red; TISSUE=ovarian follicle;
 CC RA Kobayashi D., Matsuyama M., Tanaka M., Fukada S., Nagahama Y.;
 CC RT "Structural analysis of medaka P-450c17 and expression in the ovarian
 CC RT follicle.";
 CC RL Submitted (Aug-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: CONVERSION OF PREGNENOLONE AND PROGESTERONE TO THEIR 17-
 CC CC ALPHA-HYDROXYLATED PRODUCTS AND SUBSEQUENTLY TO
 CC CC DEHYDROEPIANDROSTERONE (DHEA) AND ANDROSTENEDIONE. CATALYZES BOTH
 CC CC THE 17-ALPHA-HYDROXYLATION AND THE 17,20-LYASE REACTION.
 CC -1- CATALYTIC ACTIVITY: A steroid + AH(2) + O(2) = a 17-alpha-

```
CC hydroxysteroid + A + H(2)O.
CC -1- PATHWAY: Steroidogenic pathway; key step.
CC -1- ALTERNATIVE PRODUCTS:
CC Name=C17L;
CC Event=Alternative splicing; Named isoforms=2;
CC IsoId=P70085-1; Sequence=Displayed;
CC Name=C17S;
CC IsoId=P70085-2; Sequence=VSP_000615;
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D87121; BAA13252.1; -
CC EMBL; D87122; BAA13253.1; -
CC HSSP; P00179; 1D76.
CC InterPro: IPR001138; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Electon transport; Oxidoreductase; Monooxygenase; Membrane;
CC Heme; Steroidogenesis; Alternative splicing.
CC METL 451 451 MISSING (in isoform C17S).
CC VARSPUBLIC 230 258 /FTId=VSP_000615.
CC SEQUENCE 517 AA; 57524 MW; 7A67509437ED6C6B CRC64;
SQ
Query Match 1.6%; Score 7; DB 1; Length 517;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 83 PLEANTI 89
Db 360 PLEANTI 366
RESULT 35
CITS_BACSU STANDARD; PRT; 542 AA.
ID CITS_BACSU 034427;
AC 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sensor protein cits (EC 2.7.3.-).
GN CITS.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=AC327;
RX MEDLINE=98044033; PubMed=9384377;
RX MEDLINE=97417488; PubMed=9272861;
RA Yamamoto H., Uchiyama S., Nigroho F.A., Sekiguchi J.;
RT "Cloning and sequencing of a 35.7 kb in the 70 degree-73 degree region
RT of the Bacillus subtilis genome reveal genes for a new two-component
RT system, three spore germination proteins, an iron uptake system and a
RT general stress response protein."
RT Gene 194:191-199 (1997).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RX Kunet F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RX Azevedo V., Bertolo M.G., Bessieres P., Bolotin A., Borchert S.,
RX Borries R., Bourrier L., Brans A., Brun M., Brignell S.C., Bron S.,
RX Brouillet S., Bruch C.V., Caldwell B., Capiano V., Carter N.W.,
RX Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RX Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
```

```
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech U., Harwood C.R., Henaut A.,
RA Hilbert H., Holsteppel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klier-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maul C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portetelle S., Prescott A.M.,
RA Prescan E., Puig P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha R., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeder R., Scofield F.,
RA Sekiguchi J., Sekowska A., Serr S.J., Serr P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takouchi M., Takakoshi A., Tanaka T., Tarpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RT Nature 390:249-256 (1997).
RN (3)
RP CHARACTERIZATION.
RX STRAIN=168.
RX MEDLINE=20430113; PubMed=10972810;
RX Yamamoto H., Murata M., Sekiguchi J.;
RT "The cits two-component system regulates the expression of the
RT Mg-citrate transporter in Bacillus subtilis."
RT Mol. Microbiol. 37:898-912 (2000).
CC -1- FUNCTION: Member of the two-component regulatory system cits/cits.
CC Regulates the expression of the cits-yln operon. Functions
CC probably as a membrane-associated protein kinase that
CC phosphorylates cits in response to environmental citrate or
CC Mg(2+)-citrate complex.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -1- SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.
CC -1- SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D86417; BAA22311.1; -
CC EMBL; Z99108; CAB12587.1; -
CC PIR; E69600; E69600.
CC Subtilist; BGI2576; cits.
CC InterPro; IPR003594; ATPbind ATPase.
CC InterPro; IPR004358; Bact_sens_pr_C.
CC InterPro; IPR005467; His_Kinase.
CC InterPro; IPR000014; PAS_domain.
CC Pfam; PF02518; HAIPase_C; 1.
CC PRINTS; PR00344; BCTRLSENSOR.
CC SMART; SM00387; HAIPase_C; 1.
CC SMART; SM00091; PAS; 1.
CC PROSITE; PS50109; HIS_KIN; 1.
CC PROSITE; PS50112; PAS_FALSE_NEG.
CC Sensory transduction; Transferrase; Kinase; Transmembrane;
CC Phosphorylation; Complete proteome.
KW DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34 POTENTIAL.
FT DOMAIN 35 175 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 176 196 POTENTIAL.
FT DOMAIN 197 542 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 216 279 PAS.
FT DOMAIN 336 528 HISTIDINE KINASE.
```

FT MOD RES 339 339 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 SQ SEQUENCE 542 AA; 58891 MW; E0F3BBB9792F8C42 CRC64;
 Query Match 1.6%; Score 7; DB 1; Length 542;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 TELKXLI 238
 Db 315 TELKXLI 321

RESULT 36
 YE1F SCHPO STANDARD; PRT; 628 AA.
 AC 013879;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative transporter C1B3.15C.
 GN SPAC1B3.15C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 NX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Williams R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras K., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Gymnopoulos B.,
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Burnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong U., Forsburg S.L.,
 RA Shpakovski G.V., Ussery D., McCombie W.R., Paulsen I., Potashkin J.,
 RA "The genome sequence of Schizosaccharomyces pombe".
 RT Nature 415:871-880(2002).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: Belongs to the allantoate permease family.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z98598; CAB1241.1;
 DR PIR: T38034; T38034.
 DR GeneDB_Spombe; SPAC1B3.15C;
 KW Hypothetical protein; Transmembrane; Transport.
 FT TRANSMEM 157 177 POTENTIAL.
 FT TRANSMEM 203 223 POTENTIAL.

FT TRANSMEM 230 250 POTENTIAL.
 FT TRANSMEM 259 279 POTENTIAL.
 FT TRANSMEM 294 314 POTENTIAL.
 FT TRANSMEM 324 344 POTENTIAL.
 FT TRANSMEM 417 437 POTENTIAL.
 FT TRANSMEM 454 474 POTENTIAL.
 FT TRANSMEM 483 503 POTENTIAL.
 FT TRANSMEM 511 531 POTENTIAL.
 FT TRANSMEM 542 562 POTENTIAL.
 FT TRANSMEM 583 603 POTENTIAL.
 FT DOMAIN 92 102 POLY-SER.
 FT DOMAIN 385 388 POLY-GLU.
 SQ SEQUENCE 628 AA; 69076 MW; E1040C37445A6D36 CRC64;
 Query Match 1.6%; Score 7; DB 1; Length 628;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 VATSLIS 163
 Db 306 VATSLIS 312

RESULT 37
 IL16 HUMAN STANDARD; PRT; 631 AA.
 ID IL16 HUMAN
 AC 014005; Q15435; Q9UP18;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Interleukin-16 precursor (IL-16) (lymphocyte chemoattractant factor) (ICF).
 GN IL16.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Periphereal blood;
 RA Kornfeld H.;
 RA Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=99145586; PubMed=9990060;
 RX Bannert N., Avocat A., Baier M., Serfling E., Kurth R.;
 RT "GA-binding protein factors, in concert with the coactivator CREB
 RT binding protein/p300, control the induction of the interleukin 16
 RT promoter in T lymphocytes".
 RL Proc. Natl. Acad. Sci. U.S.A. 96:1541-1546(1999).
 RN [3]
 RP SEQUENCE OF 241-631 FROM N.A.
 RX MEDLINE=9621141; PubMed=8609984;
 RA Bannert N., Baier M., Werner A., Kurth R.;
 RT "Interleukin-16 or not?".
 RL Nature 381:30-30(1996).
 RN [4]
 RP SEQUENCE OF 502-631 FROM N.A.
 RC TISSUE=Periphereal blood;
 RX MEDLINE=9455480; PubMed=7910967;
 RA Cruikshank W.W., Center D.M., Nisar N., Wu M., Natke B.C.,
 RA Theodore A.C., Kornfeld H.;
 RT "Molecular and functional analysis of a lymphocyte chemoattractant
 RT factor: association of biologic function with CD4 expression".
 RL Proc. Natl. Acad. Sci. U.S.A. 91:5109-5113(1994).
 RN [5]
 RP SEQUENCE OF 502-631 FROM N.A.
 RA Du Y., Du G.X., Hou L.H., Wang H.T.;
 RT "cDNA sequence of interleukin-16 cloned from human PBs".
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP STRUCTURE BY NMR OF 502-631.
 RX MEDLINE=98363210; PubMed=9699630;

RA Muehlhahn P., Zweckstetter M., Georgescu J., Cicosto C., Renner C.,
 RA Landendorfer M., Lang K., Ambrosius D., Baler M., Kurch R.,
 RA Holak T.A.;
 RT "Structure of interleukin 16 resembles a PDZ domain with an occluded
 RL peptide binding site";
 CC Nat. Struct. Biol. 5:682-686(1998).
 CC -1- FUNCTION: STIMULATES A MIGRATORY RESPONSE IN CD4+ LYMPHOCYTES,
 CC MONOCYTES, AND EOSINOPHILS. ALSO INDUCES T-LYMPHOCYTE EXPRESSION
 CC OF INTERLEUKIN 2 RECEPTOR. LIGAND FOR CD4.
 CC -1- SUBUNIT: Homotrimer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Contains 2 PDZ/DHR domains.
 CC -1- DATABASE: NAME=RED Systems; cytokine source book: IL16;
 CC WWW="http://www.rndsystems.com/asp/g_sitebuilder.asp?bodyId=210".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M90391; AAD04636.1; -
 CC EMBL; AF077011; AAD15990.1; -
 CC EMBL; S81601; AAB36371.2; -
 CC EMBL; AF053412; AAC12732.1; -
 CC PIR; I59298; I59298.
 CC PDB; 1I16; 25-MAY-99.
 CC Genew; HGNC:5980; IL16.
 CC MIM; 603035; -
 CC GO; GO:0005615; Cytoplasmic space; TAS.
 CC GO; GO:0006955; P:immune response; TAS.
 CC InterPro; IPR001478; PDZ.
 CC DR Pfam; PF00595; PDZ; 2.
 CC SMART; SM00228; PDZ; 2.
 CC PROSITE; PS50106; PDZ; 2.
 CC DR Cytokine; Chemotaxis; Repeat; 3D-structure.
 CC KW PROPEP 1 ?
 CC FT CHAIN ? 631 INTERLEUKIN-16.
 CC FT DOMAIN 411 496 PDZ 1.
 CC FT DOMAIN 533 618 PDZ 2.
 CC FT CONFLICT 104 104 E -> D (IN REF. 2).
 CC FT CONFLICT 233 233 F -> L (IN REF. 2).
 CC FT CONFLICT 241 243 LRL -> PRE (IN REF. 3).
 CC FT CONFLICT 319 319 E -> A (IN REF. 1).
 CC FT STRAND 529 537
 CC FT STRAND 538 539
 CC FT TURN 543 544
 CC FT STRAND 545 548
 CC FT STRAND 561 565
 CC FT TURN 579 580
 CC FT STRAND 583 585
 CC FT STRAND 590 590
 CC FT STRAND 591 593
 CC FT HELIX 596 604
 CC FT HELIX 605 605
 CC FT TURN 610 618
 CC FT STRAND 631 AA; 66694 MW; 2A9CCED70CC8D4AC CRC64;
 CC SQ SEQUENCE

Query Match 1.6%; Score 7; DB 1; Length 631;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ELKXIE 239
 DB 388 ELKXIE 394

RESULT 38
 YFGK CAEEL STANDARD; PRT; 664 AA.
 AC 018696;

DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable 3',5'-cyclic phosphodiesterase T04D3.3 (EC 3.1.4.17).
 GN T04D3.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Kershaw J.;
 RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP REVISIONS.
 RA Jones S.J.M.;
 RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: Nucleoside 3',5'-cyclic phosphate + H(2)O =
 CC nucleoside 5'-phosphate.
 CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
 CC FAMILY. STRONG, TO MAMMALIAN TYPE 1 CGMP PHOSPHODIESTERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z81144; CAB03289.1; -
 CC PIR; T24459; T24459.
 CC DR WormPep; T04D3.3; CEI6340.
 CC DR InterPro; IPR003607; Mec_pnshphydro.
 CC DR InterPro; IPR002073; PDase.
 CC DR Pfam; PF00233; PDase; 1.
 CC DR PRINTS; PR00387; PDIESTRASE1.
 CC DR SMART; SM00471; HDC; 1.
 CC DR PROSITE; PS00126; PDASE1; 1.
 CC KW Hypothetical protein; Hydrolase; CGMP.
 CC SQ SEQUENCE 664 AA; 74854 MW; 33436A301328201E CRC64;
 CC

Query Match 1.6%; Score 7; DB 1; Length 664;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 418 TEGVLE 424
 DB 495 TEGVLE 501

RESULT 39
 ELS BOVIN STANDARD; PRT; 747 AA.
 AC P04985; P04986; P04987; Q29421;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Elastic precursor (tropoelastin).
 GN ELN.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87194772; PubMed=3032943;
 RA Raju K., Anwar R.A.;
 RT "Primary structures of bovine elastin a, b, and c deduced from the
 RL J. Biol. Chem. 262:5755-5762(1987).
 RL

RN [2]
 RP SEQUENCE OF 1-27 FROM N.A.
 RC TISSUE-Nuchal ligament;
 RX MEDLINE=89274159; PubMed=2543440;
 RA Yeh H., Anderson N., Ornstein-Goldstein N., Bashir M.M.,
 RA Rosenbloom J.C., Abrams W.R., Indik Z., Yoon K., Mecham R.,
 RA Rosenbloom J.;
 RT "Structure of the bovine elastin gene and SI nuclease analysis of
 RT alternative splicing of elastin mRNA in the bovine nuchal ligament.";
 RL Biochemistry 28:2365-2370(1989).
 RN [3]
 RP SEQUENCE OF 1-27 FROM N.A.
 RX MEDLINE=91234332; PubMed=2031719;
 RA Manohar A., Shi W., Anwar R.A.;
 RT "Partial characterization of bovine elastin gene: comparison with the
 RT gene for human elastin.";
 RL Biochem. Cell Biol. 69:185-192(1991).
 RN [4]
 RP DISULFIDE BOND.
 RX MEDLINE=92337651; PubMed=1632791;
 RA Brown P.L., Mecham R., Tisdale C., Mecham R.P.;
 RT "The cysteine residues in the carboxy terminal domain of tropoelastin
 RT form an intrachain disulfide bond that stabilizes a loop structure
 RT and positively charged pocket.";
 RL Biochem. Biophys. Res. Commun. 186:549-555(1992).
 CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
 CC NUCAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
 CC -1- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
 CC INTO AN EXTENSIBLE 3D NETWORK.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Name=1; Synonyms=A;
 CC IsoId=P04985-1; Sequence=Displayed;
 CC Name=2; Synonyms=B;
 CC IsoId=P04985-2; Sequence=VSP_004239;
 CC Name=3; Synonyms=C;
 CC IsoId=P04985-3; Sequence=VSP_004240;
 CC -1- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL J02717; AAA30503.1; -;
 CC EMBL K03505; AAA30505.1; -;
 CC EMBL K03506; AAA30506.1; -;
 CC EMBL J02855; AAA30776.1; -;
 CC EMBL M58652; AAA03519.2; -;
 CC FIR A31865; EABO.
 CC InterPro: IPR003979; tropoelastin.
 CC PRINTS: PRO1500; TROPOELASTIN.
 CC Structural protein; Connective tissue; Repeat; Signal;
 KW Alternative splicing.
 FT SIGNAL 1 26
 FT CHAIN 27 747 ELASTIN.
 FT DISULFID 737 742
 FT MOD RES 105 105 OXIDATIVE DEAMINATION.
 FT MOD RES 109 109 OXIDATIVE DEAMINATION.
 FT MOD RES 252 252 OXIDATIVE DEAMINATION.
 FT MOD RES 271 271 OXIDATIVE DEAMINATION.
 FT MOD RES 275 275 OXIDATIVE DEAMINATION.
 FT MOD RES 324 324 OXIDATIVE DEAMINATION.
 FT MOD RES 327 327 OXIDATIVE DEAMINATION.
 FT MOD RES 400 400 OXIDATIVE DEAMINATION.
 FT MOD RES 404 404 OXIDATIVE DEAMINATION.
 FT MOD RES 407 407 OXIDATIVE DEAMINATION.
 FT MOD RES 448 448 OXIDATIVE DEAMINATION.
 FT MOD RES 489 489 OXIDATIVE DEAMINATION.

FT MOD RES 493 493 OXIDATIVE DEAMINATION.
 FT MOD RES 544 544 OXIDATIVE DEAMINATION.
 FT MOD RES 548 548 OXIDATIVE DEAMINATION.
 FT MOD RES 552 552 OXIDATIVE DEAMINATION.
 FT MOD RES 606 606 OXIDATIVE DEAMINATION.
 FT MOD RES 609 609 OXIDATIVE DEAMINATION.
 FT MOD RES 645 645 OXIDATIVE DEAMINATION.
 FT MOD RES 649 649 OXIDATIVE DEAMINATION.
 FT MOD RES 655 655 OXIDATIVE DEAMINATION.
 FT MOD RES 688 688 OXIDATIVE DEAMINATION.
 FT MOD RES 688 688 OXIDATIVE DEAMINATION.
 FT VARSPLIC 226 259 Missing (in isoform 2).
 FT VARSPLIC 226 259 /FTId=VSP_004239.
 FT VARSPLIC 226 259 Missing (in isoform 3).
 FT CONFLICT 1 3 MRS -> MAG (IN REF. 2 AND 3).
 FT CONFLICT 12 12 E -> G (IN REF. 2 AND 3).
 SQ SEQUENCE 747 AA; 64229 MW; 633C03B411643D83 CR664;
 Query Match 1.6%; Score 7; DB 1; Length 747;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 371 GGVVPGA 377
 DB 458 GGVVPGA 464
 RESULT 40
 ELS_CHICK STANDARD; PRT; 750 AA.
 AC P07916;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Elastin precursor (Tropoelastin) (Fragment).
 GN ELN.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87242320; PubMed=3593675;
 RA Bressan G.M., Argos P., Stanley K.K.;
 RT "Repeating structure of chick tropoelastin revealed by complementary
 RT DNA cloning.";
 RL Biochemistry 26:1497-1503(1987).
 RN [2]
 RP SEQUENCE OF 85-750 FROM N.A. (ISOFORM 2).
 RX MEDLINE=88309083; PubMed=2841924;
 RA Baile V.J., Foster J.A.;
 RT "Multiple chick tropoelastin mRNAs";
 RL Biochem. Biophys. Res. Commun. 154:1054-1060(1988).
 RN [3]
 RP SEQUENCE OF 457-750 FROM N.A.
 RC TISSUE-Aorta;
 RX MEDLINE=87297534; PubMed=3502711;
 RA Tokimitsu I., Tajima S., Nishikawa T., Tajima M., Fukasawa T.;
 RT "Sequence analysis of elastin cDNA from chick aorta and
 RT tissue-specific transcription of the elastin gene in developing chick
 RT embryo.";
 RL Arch. Biochem. Biophys. 256:445-461(1987).
 CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
 CC NUCAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
 CC -1- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
 CC INTO AN EXTENSIBLE 3D NETWORK.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=P07916-1; Sequence=Displayed;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compgen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2003, 09:47:57 ; Search time 41 Seconds
(without alignments)
1013.230 Million cell updates/sec

Title: US-10-069-062-7

Perfect score: 432
Sequence: 1 MSKAFSPAGKAFIAGGYLV.....DIEQTEGYLSEKPEDYIGL 432

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	2.5	451	2 S57588	phosphomevalonate
2	8	1.9	224	2 S49575	ribosomal protein
3	8	1.9	312	2 T20932	hypothetical prote
4	8	1.9	340	2 C90841	probable ATP-bind
5	8	1.9	340	2 B85699	probable ATP-bind
6	8	1.9	374	2 T06018	hypothetical prote
7	8	1.9	426	2 T38650	probable phosphome
8	8	1.9	484	2 G86475	unknown protein, 1
9	8	1.9	496	2 G90433	ABC transporter, A
10	8	1.9	505	2 G84079	sodium-dependent t
11	8	1.9	505	2 C86443	unknown protein (i
12	8	1.9	645	2 G81315	membrane bound zin
13	8	1.9	953	2 S75285	methyl-accepting c
14	8	1.9	3396	2 T22613	hypothetical prote
15	7	1.6	75	2 AC1755	hypothetical prote
16	7	1.6	94	2 D82602	hypothetical prote
17	7	1.6	98	2 A70301	ribosomal protein
18	7	1.6	99	2 B72859	AcCoF-73 protein -
19	7	1.6	99	2 T41816	AcMPV orf73 - Bom
20	7	1.6	107	2 B70650	probable emr prot
21	7	1.6	108	2 B83455	conserved hypothet
22	7	1.6	108	2 A84658	hypothetical prote
23	7	1.6	129	2 B71540	hypothetical prote
24	7	1.6	137	2 A72666	hypothetical prote
25	7	1.6	143	2 S19224	ribosomal protein
26	7	1.6	168	2 P82498	conserved hypothet
27	7	1.6	186	2 B70595	hypothetical prote
28	7	1.6	206	1 E64317	hypothetical prote
29	7	1.6	217	2 AF0906	conserved hypothet

30	7	1.6	220	2 T02068	probable transcrip
31	7	1.6	220	2 C65112	sigma cross-reacti
32	7	1.6	220	2 H91139	sigma cross-reacti
33	7	1.6	220	2 C85985	sigma cross-reacti
34	7	1.6	228	2 G71700	phosphatidate cycl
35	7	1.6	247	2 F90510	conserved hypothet
36	7	1.6	258	2 T52377	PRB5 protein (imp
37	7	1.6	263	1 PKGAF	kanamycin kinase (
38	7	1.6	263	2 A13629	taurine transport
39	7	1.6	264	1 PKSOJF	kanamycin kinase (
40	7	1.6	264	2 I40613	kanamycin resistanc
41	7	1.6	265	2 F91253	sorbitose-permease p
42	7	1.6	265	2 D86094	probable sorbitose p
43	7	1.6	265	2 T05085	hypothetical prote
44	7	1.6	265	2 AD0972	lipopolysaccharide
45	7	1.6	266	2 S50189	phosphotransferase

ALIGNMENTS

RESULT 1
S57588
phosphomevalonate kinase (EC 2.7.4.2) - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: protein YM9959.02; protein YMR220w
C:Species: *Saccharomyces cerevisiae*
C>Date: 19-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C:Accession: S57588; A39606
R:Skellton, J.; Churher, C.M.
Submitted to the EMBL Data Library, June 1995
A:Reference number: S57587
A:Accession: S57588
A:Molecule type: DNA
A:Residues: 1-451 <SKB>
A:Cross-References: EMBL:Z49939; NID:9887599; PID:9887601; MIPS:YMR220w
A:Experimental source: strain AB972
R:Tsay, Y.H.; Robinson, G.W.
Mol. Cell. Biol. 11, 620-631, 1991
A:Title: Cloning and characterization of ERG8, an essential gene of *Saccharomyces cerevisiae*
A:Reference number: A39606; MUID:91117228; PMID:1846667
A:Accession: A39606
A:Molecule type: DNA
A:Residues: 1-212, 'R', 214-417, 'PLMT', 422, 'D', 424 <TSA>
A:Cross-References: EMBL:M63648; NID:9553127; PIDN:AAA34596.1; PID:g171479
C:Genetics:
A:Gene: SGD:ERG8
A:Cross-References: SGD:S0004833; MIPS:YMR220w
C:Map position: 13R
C:Keywords: phosphotransferase

Query Match 2.5%; Score 11; DB 2; Length 451;
Best Local Similarity 100.0%; Pred. No. 0.0077;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 KTGLSSAGLV 154
DB 152 KTGLSSAGLV 162

RESULT 2
S49575
ribosomal protein L10-e, cytosolic - rice
C:Species: *Oryza sativa* (rice)
C>Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 18-Aug-2000
R:Kim, J.K.
Submitted to the EMBL Data Library, September 1994
A:Description: Isolation and characterization of two rice genes encoding a putative tumo
A:Reference number: S49575
A:Accession: S49575
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-224 <KIM>

A/Cross-references: EMBL:X81691; NID:9575354; PIDN:CAA57339.1; PID:9575355
C/Genetics:
A/Gene: SC34
C/Superfamily: rat ribosomal protein L10
C/Keywords: cytosol; protein biosynthesis; ribosome

Query Match 1.9%; Score 8; DB 2; Length 224;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 APGKAFIA 14
|||||
Db 210 APGKAFIA 217

RESULT 3
T20932
hypothetical protein F14H8.2 - *Caenorhabditis elegans*
C/Species: *Caenorhabditis elegans*
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T20932

R/Mattews, L. submitted to the EMBL Data Library, October 1996
A/Reference number: Z19348
A/Accession: T20932
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-312 <MTL>

A/Cross-references: EMBL:Z81061; PIDN:CAE02931.1; GSPDB:GN00023; CESP:F14H8.2
A/Experimental source: clone F14H8
C/Genetics:
A/Gene: CESP:F14H8.2
A/Map position: 5
A/Intons: 79/3; 136/3

Query Match 1.9%; Score 8; DB 2; Length 312;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ELKKIES 240
|||||
Db 73 ELKKIES 80

RESULT 4

probable ATP-binding protein of ABC transporter Ecs1699 [imported] - *Escherichia coli* (B
C/Species: *Escherichia coli*
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C/Accession: C90841

R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gaasara, N.; Yasunaga, T.; Kuhara, S.; Shida, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene
A/Reference number: A99629; MUID:21156231; PMID:11258796
A/Accession: C90841

A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-340 <HAY>
A/Cross-references: GB:BA000007; PIDN:BAE35122.1; PID:G13361164; GSPDB:GN00154
A/Experimental source: strain O157:H7, substrain RMD 0509952
C/Genetics:
A/Gene: Ecs1699

Query Match 1.9%; Score 8; DB 2; Length 340;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 LHNVAQIA 184
|||||
Db 219 LHNVAQIA 226

RESULT 5

B85699
probable ATP-binding protein of ABC transporter Z1966 [similarity] - *Escherichia coli* (E
C/Species: *Escherichia coli*
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Nov-2001
C/Accession: B85699

R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Liao, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A/Reference number: AB5480; MUID:21074935; PMID:11206551
A/Accession: B85699
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-340 <STO>
A/Cross-references: GB:AE005174; NID:G12514906; PIDN:AA56054.1; GSPDB:GN00145; UNCP:Z19
A/Experimental source: strain O157:H7, substrain EDJ933
C/Genetics:
A/Gene: Z1966

Query Match 1.9%; Score 8; DB 2; Length 340;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 LHNVAQIA 184
|||||
Db 219 LHNVAQIA 226

RESULT 6

T06018
hypothetical protein T25K17.150 - *Arabidopsis thaliana*

C/Species: *Arabidopsis thaliana* (mouse-ear cress)
C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
C/Accession: T06018
R/Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X
submitted to the Protein Sequence Database, March 1999

A/Reference number: Z15382
A/Accession: T06018
A/Molecule type: DNA
A/Residues: 1-374 <BEV>
A/Cross-references: EMBL:AL049171; GSPDB:GN00062; ATSP:T25K17.150
A/Experimental source: cultivar Columbia; BAC clone T25K17
C/Genetics:
A/Gene: ATSP:T25K17.150
A/Map position: 4
A/Intons: 279/3

Query Match 1.9%; Score 8; DB 2; Length 374;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 VVATSLIS 163
|||||
Db 24 VVATSLIS 31

RESULT 7

T38650
probable phosphomevalonate kinase - fission yeast (*Schizosaccharomyces pombe*)
C/Species: *Schizosaccharomyces pombe*
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C/Accession: T38650

R/Murphy, L.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, August 1997

A/Reference number: Z21804
A/Accession: T38650
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-426 <MUR>
A/Cross-references: EMBL:AL109739; PIDN:CAE52264.1; GSPDB:GN00066; SPDB:SPAC343.01C
A/Experimental source: strain 972h-; cosmid c343
C/Genetics:

A:Gene: SPDB:SPAC343.01c
A:Map position: 1
A:introns: 3/2

Query Match 1.9%; Score 8; DB 2; Length 426;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 144 KTGLGSSA 151
|||||
Db 153 KTGLGSSA 160

RESULT 8

G86475

unknown protein, 13877-17163 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: G86475

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Li, J.H.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

A:Authors: Salzberrg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G86475

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-484 <STO>

A:Cross-references: GB:AB005172; NID:g10092189; PIDN:AA612607.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 1.9%; Score 8; DB 2; Length 484;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 281 EXPESSV 288
|||||
Db 402 EXPESSV 409

RESULT 9

G90433

ABC transporter, ABC1 family, probable SSO2605 [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C:Accession: G90433

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aweez, M.J.; Chan, J.; Jettifles, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.

A:Authors: Ragan, M.A.; Sensen, C.W.; Van der Oost, J.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A89139

A:Accession: G90433

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-496 <KUR>

A:Cross-references: GB:AB006641; NID:g13815908; PIDN:AAK42726.1; GSPDB:GN00155

C:Genetics:

A:Gene: SSO2605

Query Match 1.9%; Score 8; DB 2; Length 496;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 PETTYKEL 322
|||||

Db 70 PETTYKEL 77

RESULT 10

G84079

sodium-dependent transporter BH3439 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: G84079

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: G84079

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-505 <STO>

A:Cross-references: GB:AB001518; GB:BA000004; NID:g10175792; PIDN:BA07158.1; GSPDB:GN0

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH3439

C:Superfamily: gamma-aminobutyric acid transporter

Query Match 1.9%; Score 8; DB 2; Length 505;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 151 AGLVSVVA 158
|||||
Db 393 AGLVSVVA 400

RESULT 11

G86443

unknown protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: G86443

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

A:Authors: Salzberrg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G86443

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-505 <STO>

A:Cross-references: GB:AB005172; NID:g11136726; PIDN:AA631307.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 1.9%; Score 8; DB 2; Length 505;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 144 KTGLGSSA 151
|||||
Db 179 KTGLGSSA 186

RESULT 12

G81315

membrane bound zinc metalloproteinase (EC 3.4.24.-) Cj1116c [imported] - Campylobacter je

C:Species: Campylobacter jejuni

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002

C:Accession: G81315

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrel Nature 403, 665-668, 2000
 A>Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyf
 A:Reference number: AB1250; MUID:20150912; PMID:10688204
 A:Accession: G81315
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-645 <PAR>
 A:Cross-references: GB:AL139077; GB:AL111168; NID:G6968444; PIDN:CA873371.1; PID:G696855
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: ftsH; C1116c
 C:Superfamily: cell division protein ftsH; ftsH/SEC18/CDC48-type ATP-binding domain hom
 C:Keywords: hydrolase; metalloproteinase

Query Match 1.9%; Score 8; DB 2; Length 645;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 EIKKLIIES 240
 |||||
 Db 64 EIKKLIIES 71

RESULT 13
 S75285
 methyl-accepting chemotaxis protein homolog s111294 - *Synechocystis* sp. (strain PCC 6803
 C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C:Accession: S75285
 R:Kaneko, T.; Sato, S.; Kohani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A>Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
 B.

A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S75285
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-953 <KAN>
 A:Cross-references: EMBL:D90904; GB:AB001339; NID:G1652225; PIDN:BA117199.1; PID:dl01793
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics:
 A:Gene: pilJ
 C:Keywords: chemotaxis

Query Match 1.9%; Score 8; DB 2; Length 953;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 EPLTVAIK 334
 |||||
 Db 491 EPLTVAIK 498

RESULT 14
 T22613
 hypothetical protein F54B3.1 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 R:McMurray, A.
 submitted to the EMBL Data Library, March 1995
 A:Reference number: Z19589
 A:Accession: T22613
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3396 <MLT>
 A:Cross-references: EMBL:Z48583; PIDN:CA488472.1; GSPDB:GN00020; CESP:F54B3.1
 A:Experimental source: clone F54B3
 C:Genetics:
 A:Gene: CESP:F54B3.1

A:Map position: 2
 A:Introns: 52/3; 109/3; 269/1; 295/1; 432/1; 560/2; 747/2; 936/3; 1073/1; 1719/2; 1877/2
 C:Superfamily: *Caenorhabditis elegans* hypothetical protein F54B3.1

Query Match 1.9%; Score 8; DB 2; Length 3396;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 QIAGCYAQ 189
 |||||
 Db 1216 QIAGCYAQ 1223

RESULT 15
 AC1755
 hypothetical protein lin2584 [imported] - *Listeria innocua* (strain C1p11262)
 C:Species: *Listeria innocua*
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AC1755
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bioecker
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.
 D.; Jones, L.M.; Karsel, U.
 Science 294, 849-852, 2001
 A:Authors: Krefl, J.; Kuhn, M.; Kunst, F.; Kurapkac, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlund,
 A>Title: Comparative genomics of *Listeria* species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AC1755
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-75 <GLA>
 A:Cross-references: GB:AL592022; PIDN:CA97811.1; PID:G16415106; GSPDB:GN00178
 A:Experimental source: strain C1p11262
 C:Genetics:
 A:Gene: lin2584

Query Match 1.6%; Score 7; DB 2; Length 75;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 EIKKLIIE 239
 |||||
 Db 57 EIKKLIIE 63

RESULT 16
 D82602
 hypothetical protein XF2069 [imported] - *Xylella fastidiosa* (strain 9a5c)
 C:Species: *Xylella fastidiosa*
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: D82602
 R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A>Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
 A:Reference number: AB2515; MUID:20365717; PMID:10910347
 A>Note: for a complete list of authors see reference number A59328 below
 A:Accession: D82602
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-94 <SIM>
 A:Cross-references: GB:AE004023; GB:AE003849; NID:G9107185; PIDN:AAF84868.1; GSPDB:GN001
 A:Experimental source: strain 9a5c

R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Britones, W.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H
 ae-Neco, E.; Docena, C.; El-Dorri, H.; Faciniani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferrro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laizir
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmiti, D.A
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF2069

Query Match 1.6%; Score 7; DB 2; Length 94;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

27 YVTALSS 33
|||||
Db 63 YVTALSS 69

RESULT 17
A70301
ribosomal protein L22 - Aquifex aeolicus

C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 13-Aug-1999
C:Accession: A70301

R.Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O
V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: A70301

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-98 <AOP>

A:Cross-references: GB:AEO00669; NID:g2982762; PIDN:AAC06396.1; PID:g2982770; GB:AEO0065

A:Experimental source: strain VFS

C:Genetics:

A:Gene: rplV

C:Superfamily: Escherichia coli ribosomal protein L22

Query Match 1.6%; Score 7; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

421 VLEEKPE 427
|||||
Db 87 VLEEKPE 93

RESULT 18

B72859

AcOrt-73 protein - Autographa californica nuclear polyhedrosis virus

C:Species: Autographa californica nuclear polyhedrosis virus, AcMNPV

A:Note: dsDNA virus

C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 12-Nov-1999

C:Accession: B72859

R.Ayres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.

Virology 202, 586-605, 1994

A:Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.

A:Reference number: A72850; MUID:94303173; PMID:8030224

A:Accession: B72859

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-99 <AR>

A:Cross-references: GB:L22858; NID:G510708; PIDN:AAA66703.1; PID:G559142

C:Genetics:

A:Gene: AcOrt-73

Query Match 1.6%; Score 7; DB 2; Length 99;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

235 KKLIESN 241
|||||
Db 57 KKLIESN 63

RESULT 19

T41816

AcMNPV orf73 - Bombyx mori nuclear polyhedrosis virus (isolate T3)

C:Species: Bombyx mori nuclear polyhedrosis virus, BMSNV

A:Variety: isolate T3

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 11-May-2000

C:Accession: T41816

R.Gomi, S.; Majima, K.; Maeda, S.

J. Gen. Virol. 80, 1323-1337, 1999

A:Title: Sequence analysis of the genome of Bombyx mori nucleopolyhedrovirus.

A:Reference number: Z22020; MUID:99281911; PMID:10355780

A:Accession: T41816

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-99 <KAM>

A:Cross-references: EMBL:L33180; PIDN:AAC63745.1

A:Experimental source: isolate T3

C:Genetics:

A:Note: Orf_59

Query Match 1.6%; Score 7; DB 2; Length 99;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

235 KKLIESN 241
|||||
Db 57 KKLIESN 63

RESULT 20

B70650

Probable emeE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: B70650

R.Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Comor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: B70650

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-107 <COL>

A:Cross-references: GB:Z83866; GB:AL123456; NID:G3261691; PIDN:CAB06258.1; PID:G1781149

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: emeE

C:Superfamily: sngE protein

Query Match 1.6%; Score 7; DB 2; Length 107;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

156 VVATSL 162
|||||
Db 14 VVATSL 20

RESULT 21

B83455

conserved hypothetical protein PA1533 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: B83455

R.Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lim,

Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337; PMID:10984043

A/Accession: B83455
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-108 <STO>
A/Cross-references: GB:AE004581; GB:AE004091; NID:g9947482; PIDN:AA604922.1; GSPDB:GN001
A/Experimental source: strain PA01
A/Genetics: PA1533
C/Superfamily: Escherichia coli ybaB protein

Query Match 1.6%; Score 7; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 AGLVSVV 157
|||||
DB 38 AGLVSVV 44

RESULT 22

A84658
hypothetical protein AT2g26240 [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C/Accession: A84658
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentol, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.B.; Umayam, L.; Tallon, L.;
Natus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A84420; MUID:20083487; PMID:10617197
A/Accession: A84658
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-108 <STO>
A/Cross-references: GB:AE002093; NID:g3075402; PIDN:AA14534.1; GSPDB:GN00139
A/Genetics: AT2g26240
A/Map position: 2

Query Match 1.6%; Score 7; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 AGLVSVV 157
|||||
DB 86 AGLVSVV 92

RESULT 23

B71540
hypothetical protein CT222 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C/Species: Chlamydia trachomatis
C/Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C/Accession: B71540
R/Stephens, R.S.; Kaiman, S.; Lammell, C.J.; Fan, J.; Marache, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A/Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
A/Reference number: A71570; MUID:39000809; PMID:9784136
A/Accession: B71540
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1129 <ARN>
A/Cross-references: GB:AE001296; GB:AE001273; NID:g3328630; PIDN:AA67814.1; PID:g332863
A/Experimental source: serotype D, strain UW-3/Cx
A/Genetics: CT222

Query Match 1.6%; Score 7; DB 2; Length 129;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 LGSSAGL 153

|||||
DB 63 LGSSAGL 69

RESULT 24

A72666
hypothetical protein APE0752 - Aeropyrum pernix (strain K1)
C/Species: Aeropyrum pernix
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C/Accession: A72666
R/Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hatakeyama, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A/Reference number: A72450; MUID:99310339; PMID:10382966
A/Accession: A72666
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-137 <KAW>
A/Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAV9729.1; PID:d1043515; PID:g510
A/Experimental source: strain K1
C/Genetics: APE0752
C/Superfamily: Aeropyrum pernix hypothetical protein APE0752

Query Match 1.6%; Score 7; DB 2; Length 137;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 ATSLSLH 164
|||||
DB 127 ATSLSLH 133

RESULT 25

S19224
ribosomal protein L10.e, cytosolic - rice (fragment)
C/Species: Oryza sativa (rice)
C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Aug-2000
C/Accession: S19224
R/Kim, J.K.
submitted to the EMBL Data Library, February 1992
A/Description: A rice novel gene.
A/Reference number: S19224
A/Accession: S19224
A/Molecule type: mRNA
A/Residues: 1-143 <KIM>
A/Cross-references: EMBL:X64621; NID:g20311; PIDN:CAA45905.1; PID:g20312
C/Superfamily: rat ribosomal protein L10
C/Keywords: cytosol; protein biosynthesis; ribosome

Query Match 1.6%; Score 7; DB 2; Length 143;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 AFGKAPL 13
|||||
DB 134 AFGKAPL 140

RESULT 26

F82498
conserved hypothetical protein VCA0107 [imported] - Vibrio cholerae (strain N16961 serog
C/Species: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Dec-2002
C/Accession: F82498
R/Heldelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, P
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A/Reference number: A82035; MUID:20406833; PMID:10952301
A/Accession: F82498

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-168 <HEI>
A:Cross-references: GB:AE004353; GB:AE003853; NID:9657488; PIDN:AAF96021.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0107
A:Map position: 2
C:Superfamily: uncharacterized conserved protein

Query Match 1.6%; Score 7; DB 2; Length 168;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 EKKKLE 239
|||||
Db 112 EKKKLE 118

RESULT 27
B70595
hypothetical protein RV3209 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70595
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295387; PMID:9634230
A:Accession: B70595
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-186 <COL>
A:Cross-references: GB:Z95120; GB:AL123456; NID:93261739; PIDN:CAM80307.1; PID:e314511;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3209

Query Match 1.6%; Score 7; DB 2; Length 186;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 SVVATSL 161
|||||
Db 150 SVVATSL 156

RESULT 28
B64317
hypothetical protein MJ0140 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: B64317
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Teon, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, W.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A>Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: B64317
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-206 <BUN>
A:Cross-references: GB:U67471; GB:L77117; NID:92826249; PIDN:AAB98123.1; PID:g1590903; T
A:Map position: FOR138922-139542
A:Start codon: TTG
C:Superfamily: Aquifex aeolicus siroheme synthase

Query Match 1.6%; Score 7; DB 1; Length 206;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 EKKKLE 239
|||||
Db 193 EKKKLE 199

RESULT 29
AF0906
conserved hypothetical protein STY3506 [imported] - Salmonella enterica subsp. enterica
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A>Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AF0906
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AF0906
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-217 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07844.1; PID:g16504352; GSPDB:GN00176
C:Genetics:
A:Gene: STY3506
C:Superfamily: sigma cross-reacting protein 27A

Query Match 1.6%; Score 7; DB 2; Length 217;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 270 KLVSRVL 276
|||||
Db 207 KLVSRVL 213

RESULT 30
T02068
probable transcription factor QM - maize
N:Alternate names: QM protein
C:Species: Zea mays (maize)
C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 18-Aug-2000
C:Accession: T02068
R:Farmer, A.A.; Loftus, T.M.; Miller, A.A.; Sato, K.Y.; Neill, J.D.; Tron, T.; Yang, M.; Hum. Mol. Genet. 3, 723-728, 1994
A>Title: Extreme evolutionary conservation of QM, a novel c-Jun associated transcription
A:Reference number: Z14538; MUID:94362669; PMID:8081358
A:Accession: T02068
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-220 <NEI>
A:Cross-references: EMBL:U06108; NID:9468055; PIDN:AAA17419.1; PID:g468056
A:Experimental source: strain B73
C:Function:
A:Description: may act as transcription regulatory protein
A>Note: binds c-Jun and DNA
C:Superfamily: rat ribosomal protein L10
C:Keywords: DNA binding

Query Match 1.6%; Score 7; DB 2; Length 220;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 AFGKAPL 13
|||||
Db 210 AFGKAPL 216

RESULT 31
C65112
Sigma cross-reacting protein 27A - Escherichia coli (strain K-12)
C/Species: Escherichia coli
C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C/Accession: C65112; JN0287
R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; PMID:97426617; PMID:9278503
A/Accession: C65112
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-220 <HAY>
A/Cross-references: GB:AE000400; GB:U00096; NID:q2367203; PIDN:AACT6241.1; PID:g1789602;
A/Experimental source: strain K-12, substrain MG1655
R/Ishihama, R.; Fujita, N.; Ishihama, A.
Biochem. Biophys. Res. Commun. 184, 634-639, 1992
A/Title: Identification of Escherichia coli proteins cross-reacting with antibodies aga
A/Reference number: JN0286; PMID:92246944; PMID:1575737
A/Accession: JN0287
A/Molecule type: protein
A/Residues: 4-13, 'X', 15-26, 'I' <UES>
A/Comment: This protein reacts with antiserum against a synthetic peptide of RNA polymer
C/Genetics:
A/Gene: yhbL
C/Superfamily: sigma cross-reacting protein 27A

Query Match 1.6%; Score 7; DB 2; Length 220;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 270 KLVSRVL 276
DB 210 KLVSRVL 216

RESULT 32
H91139
Sigma cross-reacting protein 27A [imported] - Escherichia coli (strain O157:H7, substrain
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C/Accession: H91139
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasekara, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A/Reference number: A99629; PMID:1156231; PMID:1158796
A/Accession: H91139
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-220 <HAY>
A/Cross-references: GB:BA000007; PIDN:BA037511.1; PID:g13363561; GSPDB:GN00154
A/Experimental source: strain O157:H7, substrain RIMD 0509952
C/Genetics:
A/Gene: ECs408
C/Superfamily: sigma cross-reacting protein 27A

Query Match 1.6%; Score 7; DB 2; Length 220;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 270 KLVSRVL 276
DB 210 KLVSRVL 216

RESULT 33
CB5985
Sigma cross-reacting protein 27A (SCR-27A) [imported] - Escherichia coli (strain O157:H
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C/Accession: CB5985
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaeser, J.D.; Rose, D.J.; Mayhew
Hiller, L.; Grobeck, E.J.; Davis, N.W.; Llm, A.; Dinalanta, E.; Potamousts, K.; Apodaca,
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: AB5480; PMID:121074935; PMID:11206551
A/Accession: CB5985
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-220 <STO>
A/Cross-references: GB:AE005174; NID:g12517827; PIDN:AA658343.1; GSPDB:GN00145; UNCP:245
A/Experimental source: strain O157:H7, substrain EDJ933
C/Genetics:
A/Gene: yhbL
C/Superfamily: sigma cross-reacting protein 27A

Query Match 1.6%; Score 7; DB 2; Length 220;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 270 KLVSRVL 276
DB 210 KLVSRVL 216

RESULT 34
G71700
phosphatidate cytidyltransferase (cdaA) RP424 - Rickettsia prowazekii.
C/Species: Rickettsia prowazekii
C/Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C/Accession: G71700
R/Andersson, S.G.E.; Zomrodipour, A.; Andersson, J.O.; Stcheritz-Ponten, T.; Alsmark, U
Nature 396, 133-140, 1998
A/Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A/Reference number: A71630; PMID:99039499; PMID:9822893
A/Accession: G71700
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-228 <AND>
A/Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CA14881.1; PID:g386098
A/Experimental source: strain Madrid E
C/Genetics:
A/Gene: cdaA; RP424

Query Match 1.6%; Score 7; DB 2; Length 228;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 SAGLVSV 156
DB 140 SAGLVSV 146

RESULT 35
F90510
conserved hypothetical protein [imported] - Sulfolobus solfataricus
C/Species: Sulfolobus solfataricus
C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C/Accession: F90510
R/She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A/Description: Sulfolobus solfataricus complete genome.
A/Reference number: A99139
A/Accession: F90510
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-247 <KUP>
A/Cross-references: GB:AE006641; NID:g13816696; PIDN:AAK43343.1; GSPDB:GN00155
C/Genetics:
A/Gene: SSO3250

Query Match 1.6%; Score 7; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 361 DAIATLV 387
46 DAIATLV 52

RESULT 36

T52377
PRIB5 protein [imported] - *Ribes nigrum* (fragment)

C:Species: *Ribes nigrum*
C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 03-Nov-2000
C:Accession: T52377
R:Woodhead, M.R.; Taylor, M.A.; Brennan, R.M.; McNicol, R.J.; Davies, H.V.
submitted to the EMBL Data Library, June 1998
A:Description: Cloning and characterisation of the cDNA clones of five genes that are di
A:Reference number: Z26059
A:Accession: T52377
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-258 <MO>
A:Cross-references: EMBL:AJ007578; PIDN:CAA07566.1
A:Experimental source: cultivar ben alder
C:Genetics:
A:Gene: PRIB5
C:Superfamily: Arabidopsis thaliana hypothetical protein T8B10.110

Query Match 1.6%; Score 7; DB 2; Length 258;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 323 DHSVEPL 329
Db 150 DHSVEPL 156

RESULT 37

PKSAP

kanamycin kinase (EC 2.7.1.95) - *Staphylococcus aureus*
N:Alternate names: aminoglycoside 3'-phosphotransferase; neomycin-kanamycin phosphotrans
C:Species: *Staphylococcus aureus*
C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 13-Sep-1997
C:Accession: A26220; A00665
R:Gray, G.S.; Filch, W.M.
Mol. Biol. Evol. 1, 57-66, 1983
A:Title: Evolution of antibiotic resistance genes: the DNA sequence of a kanamycin resis
A:Reference number: A93054; MUID:88174299; PMID:6100986
A:Accession: A26220
A:Molecule type: DNA
A:Residues: 1-263 <GRA>
C:Superfamily: Kanamycin kinase
C:Keywords: antibiotic resistance; ATP; phosphotransferase
F:189/Active site: Asp #status predicted

Query Match 1.6%; Score 7; DB 1; Length 263;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 ELKKLIE 239
Db 9 ELKKLIE 15

RESULT 38

A13629
taurine transport permease protein tauc [imported] - *Brucella melitensis* (strain 16M)
C:Species: *Brucella melitensis*
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: A13629
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Leless

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
A:Reference number: AD3252; PMID:11756688

A:Accession: A13629
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-263 <KDR>
A:Cross-references: GB:AE008918; PIDN:AAU54204.1; PID:G17985173; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI10962
A:Map position: 11

Query Match 1.6%; Score 7; DB 2; Length 263;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 147 LGSSAGL 153
Db 203 LGSSAGL 209

RESULT 39

PKSOP

kanamycin kinase (EC 2.7.1.95) - *Enterococcus faecalis* plasmid pPH1
N:Alternate names: aminoglycoside 3'-phosphotransferase; neomycin-kanamycin phosphotrans
C:Species: *Enterococcus faecalis*
C:Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 24-Sep-1999
C:Accession: A00665
R:Ritieu-Cuot, P.; Courvalin, P.
Gene 23, 331-341, 1983
A:Title: Nucleotide sequence of the *Streptococcus faecalis* plasmid gene encoding the 3'
A:Reference number: A00665; MUID:84029883; PMID:6313476
A:Accession: A00665
A:Molecule type: DNA
A:Residues: 1-264 <TRI>
A:Cross-references: GB:V01547; NID:G47033; PIDN:CAA24789.1; PID:G47034
C:Genetics:
A:Genome: plasmid
C:Superfamily: Kanamycin kinase
C:Keywords: antibiotic resistance; ATP; phosphotransferase
F:190/Active site: Asp #status predicted

Query Match 1.6%; Score 7; DB 1; Length 264;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 ELKKLIE 239
Db 9 ELKKLIE 15

RESULT 40

I40613
kanamycin resistance protein - *Campylobacter coli* plasmid pPI433
C:Species: *Campylobacter coli*
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 18-Jun-1999
C:Accession: I40613
R:Ritieu-Cuot, P.; Gerbaud, G.; Lambert, T.; Courvalin, P.
EMBO J. 4, 3583-3587, 1985
A:Title: In vivo transfer of genetic information between gram-positive and gram-negative
A:Reference number: I40613; MUID:86135998; PMID:3937729
A:Accession: I40613
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-264 <RBS>
A:Cross-references: GB:M26832; NID:G144184; PIDN:AAA98050.1; PID:G144185
A:Experimental source: plasmid pPI433
C:Genetics:
A:Gene: aphA-3
A:Genome: plasmid
C:Superfamily: Kanamycin kinase
C:Keywords: ATP

Query Match 1.6%; Score 7; DB 2; Length 264;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ELKLE 239
 |||||
 DB 9 ELKLE 15

Search completed: August 8, 2003, 09:53:27
 Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2003, 09:52:48 ; Search time 52 Seconds
(without alignments)
986.620 Million cell updates/sec

Title: US-10-069-062-7

Perfect score: 432
Sequence: 1 MSKAFAPGKAFIAGGYLV.....DIEQTEGVLEKPEHYIGL 432

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 451899 seqs, 118759770 residues

Word size : 0

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: Published Applications AA.*
2: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/PCF_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCFUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	3.0	451	US-10-128-714-3406	Sequence 3406, App
2	13	3.0	484	US-10-128-714-4406	Sequence 8406, App
3	11	2.5	451	US-09-909-745-24	Sequence 24, App1
4	8	1.9	503	US-10-036-959B-12	Sequence 12, App1
5	8	1.9	505	US-09-988-863A-2	Sequence 2, App1
6	7	1.6	86	US-09-774-381-50	Sequence 50, App1
7	7	1.6	88	US-10-156-761-13131	Sequence 13131, A
8	7	1.6	108	US-09-815-242-11831	Sequence 11831, A
9	7	1.6	108	US-10-127-032-139	Sequence 139, App
10	7	1.6	115	US-09-864-761-48749	Sequence 48749, A
11	7	1.6	146	US-09-796-692-1450	Sequence 1450, App
12	7	1.6	146	US-10-040-862-1900	Sequence 1450, App
13	7	1.6	146	US-10-040-862-1450	Sequence 1450, App
14	7	1.6	146	US-10-040-862-1900	Sequence 1900, App
15	7	1.6	149	US-09-989-920-279	Sequence 279, App

16	7	1.6	157	9	US-09-867-550-1888	Sequence 1888, App
17	7	1.6	179	15	US-10-157-031-411	Sequence 411, App
18	7	1.6	184	10	US-09-738-973-427	Sequence 427, App
19	7	1.6	184	10	US-09-854-133-427	Sequence 427, App
20	7	1.6	184	15	US-10-144-649A-427	Sequence 427, App
21	7	1.6	186	16	US-10-080-170-587	Sequence 587, App
22	7	1.6	204	9	US-09-828-644-75	Sequence 75, App1
23	7	1.6	258	9	US-09-800-528-6	Sequence 6, App1
24	7	1.6	263	10	US-09-738-626-5084	Sequence 5084, App
25	7	1.6	291	9	US-09-815-242-11818	Sequence 11818, A
26	7	1.6	320	10	US-09-886-055-121	Sequence 121, App
27	7	1.6	320	11	US-09-804-291-121	Sequence 121, App
28	7	1.6	360	9	US-09-864-761-46708	Sequence 46708, A
29	7	1.6	360	10	US-09-738-626-5241	Sequence 5241, App
30	7	1.6	396	9	US-09-815-242-10968	Sequence 10968, A
31	7	1.6	405	9	US-09-805-550-2	Sequence 2, App1
32	7	1.6	407	15	US-10-157-031-281	Sequence 281, App
33	7	1.6	472	10	US-09-978-295A-245	Sequence 245, App
34	7	1.6	472	10	US-09-978-697-245	Sequence 245, App
35	7	1.6	472	10	US-09-978-192A-245	Sequence 245, App
36	7	1.6	472	10	US-09-999-932A-245	Sequence 245, App
37	7	1.6	472	11	US-09-978-189-245	Sequence 245, App
38	7	1.6	472	11	US-09-978-608A-245	Sequence 245, App
39	7	1.6	472	11	US-09-978-585A-245	Sequence 245, App
40	7	1.6	472	11	US-09-978-191A-245	Sequence 245, App
41	7	1.6	472	11	US-09-978-403A-245	Sequence 245, App
42	7	1.6	472	11	US-09-978-564A-245	Sequence 245, App
43	7	1.6	472	11	US-09-999-933A-245	Sequence 245, App
44	7	1.6	472	11	US-09-981-915A-245	Sequence 245, App
45	7	1.6	472	11	US-09-978-824-245	Sequence 245, App

ALIGNMENTS

RESULT 1
US-10-128-714-3406
Sequence 3406, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Wenqi
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Eroshkin, Alexey M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10128, 714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3406
LENGTH: 451
TYPE: PRT
ORGANISM: Aspergillus fumigatus
US-10-128-714-3406
Query Match 3.0%; Score 13; DB 15; Length 451;
Best Local Similarity 100.0%; Pred. No. 0.00042; Indels 0; Gaps 0;
Matches 13; Conservative 0; Mismatches 0;

QY 371 GGVPGAGGYDAI 383
|||||
DB 379 GGVPGAGGYDAI 391

RESULT 2

US-10-128-714-8406
; Sequence 8406, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Ebroekhin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in *Aspergillus fumigatus* and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; PRIOR FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8406
; LENGTH: 484
; TYPE: PRT
; ORGANISM: *Aspergillus fumigatus*
US-10-128-714-8406

Query Match 3.0%; Score 13; DB 15; Length 484;
Best Local Similarity 100.0%; Pred. No. 0.00044;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 GGVPGAGGYDAI 383
|||||
DB 412 GGVPGAGGYDAI 424

RESULT 3

US-09-909-745-24
; Sequence 24, Application US/09909745
; Patent No. US20020119546A1
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Omolayo O.
; TITLE OF INVENTION: Squalene Synthesis Enzymes
; FILE REFERENCE: B0112 US CIP
; CURRENT APPLICATION NUMBER: US/09/909,745
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 60/107,241
; PRIOR FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 60/107,241
; PRIOR FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 451
; TYPE: PRT
; ORGANISM: *Saccharomyces cerevisiae*
US-09-909-745-24

Query Match 2.5%; Score 11; DB 10; Length 451;
Best Local Similarity 100.0%; Pred. No. 0.037;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 144 KTGLGSSAGLV 154
|||||
DB 152 KTGLGSSAGLV 162

RESULT 4

US-10-036-959B-12
; Sequence 12, Application US/10036959B
; Publication No. US20030119098A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours & Company
; APPLICANT: Hallahan, David L.
; APPLICANT: Keiper-Hyenko, Natalie
; TITLE OF INVENTION: Genes Involved in the Biosynthesis of Isopentenyl Diphosphate in
; FILE REFERENCE: CL-1792
; CURRENT APPLICATION NUMBER: US/10/036,959B
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: 60/307,637
; PRIOR FILING DATE: 2001-07-25
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 503
; TYPE: PRT
; ORGANISM: *Hevea brasiliensis*
US-10-036-959B-12

Query Match 1.9%; Score 8; DB 15; Length 503;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 KTGLGSSA 151
|||||
DB 180 KTGLGSSA 187

RESULT 5

US-09-988-863A-2
; Sequence 2, Application US/0998863A
; Patent No. US20020123427A1
; GENERAL INFORMATION:
; APPLICANT: Bayer AG
; TITLE OF INVENTION: Plant phosphomevalonate kinases
; FILE REFERENCE: Le A 35 018
; CURRENT APPLICATION NUMBER: US/09/988,863A
; CURRENT FILING DATE: 2001-11-21
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 505
; TYPE: PRT
; ORGANISM: *Arabidopsis thaliana*
US-09-988-863A-2

Query Match 1.9%; Score 8; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 KTGLGSSA 151
|||||
DB 179 KTGLGSSA 186

RESULT 6

US-09-774-381-50
; Sequence 50, Application US/09774381
; Publication No. US20030082677A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: McCarthy, Sean A.


```
APPLICANT: Pan, Yang
APPLICANT: Gearing, David P.
TITLE OF INVENTION: NOVEL EDIIRF, MTR-1, ISP-1, TAP-1, AND PA-1 MOLECULES
TITLE OF INVENTION: AND USES THEREFOR
FILE REFERENCE: MNT-107CP2
CURRENT APPLICATION NUMBER: US/09/774,381
CURRENT FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 08/941,354
PRIOR FILING DATE: 1999-09-30
PRIOR APPLICATION NUMBER: 09/010,674
PRIOR FILING DATE: 1998-01-22
PRIOR APPLICATION NUMBER: 60/061,149
PRIOR FILING DATE: 1997-10-06
PRIOR APPLICATION NUMBER: 09/014,347
PRIOR FILING DATE: 1998-01-27
PRIOR APPLICATION NUMBER: 60/061,159
PRIOR FILING DATE: 1997-10-06
PRIOR APPLICATION NUMBER: 09/474,151
PRIOR FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: 09/004,206
PRIOR FILING DATE: 1998-01-08
PRIOR APPLICATION NUMBER: 60/061,143
PRIOR FILING DATE: 1997-10-06
PRIOR APPLICATION NUMBER: 09/483,414
PRIOR FILING DATE: 2000-01-14
PRIOR APPLICATION NUMBER: 09/213,571
PRIOR FILING DATE: 1998-12-18
PRIOR APPLICATION NUMBER: 08/994,890
PRIOR FILING DATE: 1997-12-19
NUMBER OF SEQ ID NOS: 59
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 50
LENGTH: 86
TYPE: PRT
ORGANISM: Homo sapiens
US-09-774-381-50
```

```
Query Match 1.6%; Score 7; DB 11; Length 86;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 326 VEPITVA 332
    |||||
Db 34 VEPITVA 40
```

```
RESULT 7
US-10-156-761-13131
Sequence 13131, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-24089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 13131
LENGTH: 88
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-13131
```

```
Query Match 1.6%; Score 7; DB 15; Length 88;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 270 KLVSRVL 276
    |||||
Db 78 KLVSRVL 84
```

```
RESULT 8
US-09-815-242-11831
Sequence 11831, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyckind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11831
LENGTH: 108
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11831
```

```
Query Match 1.6%; Score 7; DB 9; Length 108;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 151 AGVSVV 157
    |||||
Db 38 AGVSVV 44
```

```
RESULT 9
US-10-127-032-139
Sequence 139, Application US/10127032
Publication No. US20030113742A1
GENERAL INFORMATION:
APPLICANT: Whiteley, Marvin
APPLICANT: Banger, M. Gita
APPLICANT: Lory, Stephen
APPLICANT: Greenberg, Everett Peter
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE MODULATION OF
FILE REFERENCE: UZ-070CP
CURRENT APPLICATION NUMBER: US/10/127,032
CURRENT FILING DATE: 2002-04-19
PRIOR APPLICATION NUMBER: US 60/285,190
```

PRIOR FILING DATE: 2001-04-20
PRIOR APPLICATION NUMBER: US 60/344,142
PRIOR FILING DATE: 2001-10-24
NUMBER OF SEQ ID NOS: 170
SOFTWARE: FastSeq for windows Version 4.0
SEQ ID NO 139
LENGTH: 108
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-10-127-032-139

Query Match
Best Local Similarity 1.6%; Score 7; DB 15; Length 108;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 AGLSVY 157
DB 38 AGLSVY 44

RESULT 10

US-09-864-761-48749
Sequence 48749, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME- DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 48749

LENGTH: 115
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004080.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9
OTHER INFORMATION: EST HUMAN HIT: A1346148.1, EVALUATE 9.00e-42
OTHER INFORMATION: SWISSPROT HIT: P49640, EVALUATE 2.00e-36
US-09-864-761-48749

Query Match
Best Local Similarity 1.6%; Score 7; DB 9; Length 115;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 GGSERPK 270
DB 37 GGSERPK 43

RESULT 11

US-09-796-692-1450
Sequence 1450, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
FILE REFERENCE: 2077.001200
CURRENT APPLICATION NUMBER: US/09/796,692
PRIOR FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: 60/223,378
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 9597
SOFTWARE: FastSeq for windows Version 3.0
SEQ ID NO 1450
LENGTH: 146
TYPE: PRT
ORGANISM: Homo sapiens
US-09-796-692-1450

Query Match
Best Local Similarity 1.6%; Score 7; DB 10; Length 146;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ELKKLIE 239
DB 56 ELKKLIE 62

```
RESULT 12
US-09-796-692-1900
; Sequence 1900, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1900
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-1900

Query Match      1.6%; Score 7; DB 10; Length 146;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      233 ELKULIE 239
Db      56 ELKULIE 62

RESULT 13
US-10-040-862-1450
; Sequence 1450, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
```

```
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1450
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-040-862-1450

Query Match      1.6%; Score 7; DB 15; Length 146;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      233 ELKULIE 239
Db      56 ELKULIE 62

RESULT 14
US-10-040-862-1900
; Sequence 1900, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
```

;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: US 60/223,416
;; PRIOR FILING DATE: 2000-08-04
;; PRIOR APPLICATION NUMBER: US 60/223,378
;; PRIOR FILING DATE: 2000-08-07
;; PRIOR APPLICATION NUMBER: US 09/796,692
;; PRIOR FILING DATE: 2001-03-01
;; NUMBER OF SEQ ID NOS: 10467
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 1900
;; LENGTH: 146
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-040-862-1900

Query Match
Best Local Similarity 100.0%; Score 7; DB 15; Length 146;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ELKGLIE 239
DB 56 ELKGLIE 62

RESULT 15
US-09-989-920-279
;; Sequence 279, Application US/09989920
;; Patent No. US20020172857A1
;; GENERAL INFORMATION:
;; APPLICANT: Macina, Roberto
;; APPLICANT: Recipon, Hervé
;; APPLICANT: Chen, Sei-Yu
;; APPLICANT: Sun, Yongming
;; APPLICANT: Liu, Chenghua
;; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Pro
;; FILE REFERENCE: DEX-0291
;; CURRENT APPLICATION NUMBER: US/09/989,920
;; CURRENT FILING DATE: 2001-11-21
;; PRIOR APPLICATION NUMBER: 60/252,500
;; PRIOR FILING DATE: 2000-11-22
;; NUMBER OF SEQ ID NOS: 284
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 279
;; LENGTH: 149
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-989-920-279

Query Match
Best Local Similarity 100.0%; Score 7; DB 10; Length 149;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 369 CVCGVVP 375
DB 124 CVCGVVP 130

RESULT 16
US-09-867-550-1888
;; Sequence 1888, Application US/09867550
;; Patent No. US20020082206A1
;; GENERAL INFORMATION:
;; APPLICANT: Leach, Martin D.
;; APPLICANT: Mehraban, Peadar
;; APPLICANT: Conley, Pamela
;; APPLICANT: Law, Debbie
;; APPLICANT: Topper, James
;; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and
;; TITLE OF INVENTION: Thereby
;; FILE REFERENCE: 21402-013 (Cura-313)
;; CURRENT APPLICATION NUMBER: US/09/867,550
;; CURRENT FILING DATE: 2001-09-20
;; PRIOR APPLICATION NUMBER: USSN 60/208,427

;; PRIOR FILING DATE: 2000-05-30
;; NUMBER OF SEQ ID NOS: 2125
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 1888
;; LENGTH: 157
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: VARIANT
;; LOCATION: (120)
;; OTHER INFORMATION: Wherein Xaa may be any one of Cys or Phe or Ser or Tyr
US-09-867-550-1888

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 157;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 RRFQPAL 215
DB 110 RRFQPAL 116

RESULT 17
US-10-157-031-411
;; Sequence 411, Application US/10157031
;; Publication No. US20030108890A1
;; GENERAL INFORMATION:
;; APPLICANT: Baranova, A. V.
;; APPLICANT: Yankovsky, N. K.
;; APPLICANT: Kozlov, A. P.
;; APPLICANT: Lobashev, A. V.
;; APPLICANT: Krukovskaya, L. L.
;; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
;; FILE REFERENCE: 2760-103
;; CURRENT APPLICATION NUMBER: US/10/157,031
;; CURRENT FILING DATE: 2002-05-30
;; NUMBER OF SEQ ID NOS: 415
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 411
;; LENGTH: 179
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-157-031-411

Query Match
Best Local Similarity 100.0%; Score 7; DB 15; Length 179;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 EATIFIV 92
DB 86 EATIFIV 92

RESULT 18
US-09-738-973-427
;; Sequence 427, Application US/09738973
;; Patent No. US20020110563A1
;; GENERAL INFORMATION:
;; APPLICANT: Reed, Steven G.
;; APPLICANT: Henderson, Robert A.
;; APPLICANT: Lodes, Michael J.
;; APPLICANT: Fling, Steven P.
;; APPLICANT: Mohamach, Radodh
;; APPLICANT: Algate, Paul A.
;; APPLICANT: Secrist, Heather
;; APPLICANT: Indratis, Carol Yoseph
;; APPLICANT: Benson, Darin R.
;; APPLICANT: Elliot, Mark
;; APPLICANT: Mannion, Jane
;; APPLICANT: Kalos, Michael D.
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
;; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
;; FILE REFERENCE: 210121.475C9

;; CURRENT APPLICATION NUMBER: US/09/738,973
;; CURRENT FILING DATE: 2000-12-14
;; NUMBER OF SEQ ID NOS: 587
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 427
;; LENGTH: 184
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-738-973-427

Query Match 1.6%; Score 7; DB 10; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 EATIFIV 92
Db 91 EATIFIV 97

RESULT 19
US-09-854-133-427
; Sequence 427, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raedoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 427
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-854-133-427

Query Match 1.6%; Score 7; DB 10; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 EATIFIV 92
Db 91 EATIFIV 97

RESULT 20
US-10-144-649A-427
; Sequence 427, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongrong
; APPLICANT: Fan, Liqun
; APPLICANT: Algate, Paul A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C11
; CURRENT APPLICATION NUMBER: US/10/144,649A
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 749
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 427
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-649A-427

Query Match 1.6%; Score 7; DB 15; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 EATIFIV 92
Db 91 EATIFIV 97

RESULT 21
US-10-080-170-587
; Sequence 587, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 587
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170-587

Query Match 1.6%; Score 7; DB 16; Length 186;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 SVVATSL 161
Db 150 SVVATSL 156

RESULT 22
US-09-828-644-75
; Sequence 75, Application US/09828644
; Patent No. US20020015998A1
; GENERAL INFORMATION:
; APPLICANT: Vogeli, Gabriel
; TITLE OF INVENTION: NO. US20020015998A1 G Protein-Coupled Receptors
; FILE REFERENCE: 00196US1
; CURRENT APPLICATION NUMBER: US/09/828,644
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/195,150
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,099
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,151
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,148
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,093
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,098
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/230,149
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 75
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-828-644-75

Query Match 1.6%; Score 7; DB 9; Length 204;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 FILEATI 89
DB 27 FILEATI 33

RESULT 23

US-09-800-528-6
; Sequence 6, Application US/09800528
; Patent No. US20010016955A1
; GENERAL INFORMATION:
; APPLICANT: Woodhead, Mary Rose
; APPLICANT: Brennan, Rex Michael
; TITLE OF INVENTION: BLACKCURRANT PROMOTERS AND GENES
; FILE REFERENCE: C70237D1
; CURRENT APPLICATION NUMBER: US/09/800,528
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 09/068,140
; PRIOR FILING DATE: 1998-11-02
; PRIOR APPLICATION NUMBER: PCT/EP/9604807
; PRIOR FILING DATE: 1996-11-04
; PRIOR APPLICATION NUMBER: GB9522558.7
; PRIOR FILING DATE: 1995-11-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Ribes nigrum
US-09-800-528-6

Query Match 1.6%; Score 7; DB 9; Length 258;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 DHSVEPL 329
DB 150 DHSVEPL 156

RESULT 24

US-09-738-626-5084
; Sequence 5084, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, WASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5084
; LENGTH: 263

TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5084

Query Match 1.6%; Score 7; DB 10; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 TAYGLI 206
DB 243 TAYGLI 249

RESULT 25

US-09-815-242-11818
; Sequence 11818, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11818
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11818

Query Match 1.6%; Score 7; DB 9; Length 291;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 BEKPEYD 429
DB 200 BEKPEYD 206

RESULT 26

US-09-886-055-121
; Sequence 121, Application US/09886055
; Patent No. US20020132273A1
; GENERAL INFORMATION:
; APPLICANT: STRYER, LUBERT
; APPLICANT: ZOZULYA, SERGEY
; TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND
; FILE REFERENCE: 078003-0277150
; CURRENT APPLICATION NUMBER: US/09/886,055

```

; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 60/213,812
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 522
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 121
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-886-055-121

Query Match
Best Local Similarity 1.6%; Score 7; DB 10; Length 320;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 201 AUYGLIV 207
Db 198 AUYGLIV 204

RESULT 27
US-09-804-291-121
; Sequence 121, Application US/09804291
; Publication No. US20030088059A1
; GENERAL INFORMATION:
; APPLICANT: ZOZUA, SERGEY
; TITLE OF INVENTION: HUMAN OLFACTORY RECEPTORS AND GENES ENCODING SAME
; FILE REFERENCE: P 0278005
; CURRENT APPLICATION NUMBER: US/09/804,291
; CURRENT FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/188,914
; PRIOR FILING DATE: 2000-03-13
; PRIOR APPLICATION NUMBER: 60/192,033
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/198,474
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/199,335
; PRIOR FILING DATE: 2000-04-24
; PRIOR APPLICATION NUMBER: 60/207,702
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/213,849
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/226,534
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: 60/230,732
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 60/266,862
; PRIOR FILING DATE: 2001-02-07
; NUMBER OF SEQ ID NOS: 529
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 121
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-804-291-121

Query Match
Best Local Similarity 1.6%; Score 7; DB 11; Length 320;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 201 AUYGLIV 207
Db 198 AUYGLIV 204

RESULT 28
US-09-864-761-46708
; Sequence 46708, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
```

```

; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46708
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL031276.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: Q07878, EVALU8 1.60e-01
; OTHER INFORMATION: EST_HUMAN HIT: BF362311.1, EVALU8 2.00e-54
US-09-864-761-46708

Query Match
Best Local Similarity 1.6%; Score 7; DB 9; Length 360;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 125 SSNGEKT 131
Db 39 SSNGEKT 45

RESULT 29
US-09-738-626-5241
; Sequence 5241, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
```

```
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5241
LENGTH: 360
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5241
```

```
Query Match
Best Local Similarity 1.6%; Score 7; DB 10; Length 360;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 218 DVEQVLE 224
DB 331 DVEQVLE 337
```

```
RESULT 30
US-09-815-242-10968
Sequence 10968 Application US/09815242
Patent No. US2002061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10968
LENGTH: 396
```

```
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-815-242-10968
```

```
Query Match
Best Local Similarity 1.6%; Score 7; DB 9; Length 396;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 197 DVATATAY 203
DB 332 DVATATAY 338
```

```
RESULT 31
US-09-805-550-2
Sequence 2 Application US/09805550
Patent No. US20020026045A1
GENERAL INFORMATION:
APPLICANT: Mahajan, Pramod B.
APPLICANT: Taghian, Laura
TITLE OF INVENTION: Rad23 Genes and Uses Thereof
FILE REFERENCE: 0964D
CURRENT APPLICATION NUMBER: US/09/805,550
CURRENT FILING DATE: 2001-03-13
PRIOR APPLICATION NUMBER: 09/413,574
PRIOR FILING DATE: 1999-10-06
PRIOR APPLICATION NUMBER: 60/109,728
PRIOR FILING DATE: 1998-11-23
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 405
TYPE: PRT
ORGANISM: Zea mays
US-09-805-550-2
```

```
Query Match
Best Local Similarity 1.6%; Score 7; DB 9; Length 405;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 372 GVPGAG 378
DB 266 GVPGAG 272
```

```
RESULT 32
US-10-157-031-281
Sequence 281 Application US/10157031
Publication No. US20030108890A1
GENERAL INFORMATION:
APPLICANT: Baranova, A. V.
APPLICANT: Yankovsky, N. K.
APPLICANT: Kozlov, A. P.
APPLICANT: Lobashev, A. V.
APPLICANT: Krukovskaya, L. L.
TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
FILE REFERENCE: 2760-103
CURRENT APPLICATION NUMBER: US/10/157,031
CURRENT FILING DATE: 2002-05-30
NUMBER OF SEQ ID NOS: 415
SOFTWARE: PatentIn version 3.1
SEQ ID NO 281
LENGTH: 407
TYPE: PRT
ORGANISM: Homo sapiens
US-10-157-031-281
```

```
Query Match
Best Local Similarity 1.6%; Score 7; DB 15; Length 407;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 264 GGSSTPK 270
DB 331 GGSSTPK 337
```


DB 155 GGSFTRK 161

RESULT 33
US-09-978-295A-245
Sequence 245, Application US/09978295A
Patent No. US2002015606A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi
APPLICANT: Baker Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan
APPLICANT: Ferreira, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerltsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same.
FILE REFERENCE: P2630P1C11
CURRENT APPLICATION NUMBER: US/09/978,295A
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/07450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656

PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29

;; PRIOR APPLICATION NUMBER: 60/083495
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083496
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083499
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083545
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083554
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083558
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083559
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083500
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083742
;; PRIOR FILING DATE: 1998-04-30
;; PRIOR APPLICATION NUMBER: 60/084366
;; PRIOR FILING DATE: 1998-05-05
;; PRIOR APPLICATION NUMBER: 60/084414
;; PRIOR FILING DATE: 1998-05-06
;; PRIOR APPLICATION NUMBER: 60/084441
;; PRIOR FILING DATE: 1998-05-06
;; PRIOR APPLICATION NUMBER: 60/084637
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084639
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084640
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084598
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084600
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084627
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084643
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/085339
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085338
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085323
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085582
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085700
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085689
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085579
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085580
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085573
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085704
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085697

Query Match 1.6%; Score 7; DB 10; Length 472;
Best Local Similarity 100.0%; Pred. No. 36+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 KKEKPE 285
DB 183 KKEKPE 189

RESULT 34
US-09-978-697-245
; Sequence 245, Application US/09978697
; Patent No. US20020169284A1

;; GENERAL INFORMATION:
;; APPLICANT: Ashkenazi, Avi
;; APPLICANT: Baker Kevin P.
;; APPLICANT: Botstein, David
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Eaton, Dan
;; APPLICANT: Ferrara, Napoleon
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Fong, Sherman
;; APPLICANT: Gao, Wei-Qiang
;; APPLICANT: Gerber, Hanspeter
;; APPLICANT: Gerritsen, Mary E.
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Hillan, Kenneth J.
;; APPLICANT: Kilavin, Ivar J.
;; APPLICANT: Kuo, Sophia S.
;; APPLICANT: Napier, Mary A.
;; APPLICANT: Pan, James;
;; APPLICANT: Paoni, Nicholas F.
;; APPLICANT: Roy, Margaret Ann
;; APPLICANT: Shelton, David L.
;; APPLICANT: Stewart, Timothy A.
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Williams, P. Mickey
;; APPLICANT: Wood, William I.
;; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;; TITLE OF INVENTION: Acids Encoding the Same
;; FILE REFERENCE: P2630PLC27
;; CURRENT FILING DATE: 2001-10-16
;; PRIOR APPLICATION NUMBER: 09/918585
;; PRIOR FILING DATE: 2001-07-30
;; PRIOR APPLICATION NUMBER: 60/062250
;; PRIOR FILING DATE: 1997-10-17
;; PRIOR APPLICATION NUMBER: 60/064249
;; PRIOR FILING DATE: 1997-11-03
;; PRIOR APPLICATION NUMBER: 60/065311
;; PRIOR FILING DATE: 1997-11-13
;; PRIOR APPLICATION NUMBER: 60/066364
;; PRIOR FILING DATE: 1997-11-21
;; PRIOR APPLICATION NUMBER: 60/077450
;; PRIOR FILING DATE: 1998-03-10
;; PRIOR APPLICATION NUMBER: 60/077632
;; PRIOR FILING DATE: 1998-03-11
;; PRIOR APPLICATION NUMBER: 60/077641
;; PRIOR FILING DATE: 1998-03-11
;; PRIOR APPLICATION NUMBER: 60/077649
;; PRIOR FILING DATE: 1998-03-11
;; PRIOR APPLICATION NUMBER: 60/077791
;; PRIOR FILING DATE: 1998-03-12
;; PRIOR APPLICATION NUMBER: 60/078004
;; PRIOR FILING DATE: 1998-03-13
;; PRIOR APPLICATION NUMBER: 60/078886
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/078936
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/078910
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/078939
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/079294
;; PRIOR FILING DATE: 1998-03-25
;; PRIOR APPLICATION NUMBER: 60/079656
;; PRIOR FILING DATE: 1998-03-26
;; PRIOR APPLICATION NUMBER: 60/079664
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/079663
;; PRIOR FILING DATE: 1998-03-27

;; PRIOR APPLICATION NUMBER: 60/079728
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/079786
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/079920
;; PRIOR FILING DATE: 1998-03-30
;; PRIOR APPLICATION NUMBER: 60/079923
;; PRIOR FILING DATE: 1998-03-30
;; PRIOR APPLICATION NUMBER: 60/080105
;; PRIOR FILING DATE: 1998-03-31
;; PRIOR APPLICATION NUMBER: 60/080107
;; PRIOR FILING DATE: 1998-03-31
;; PRIOR APPLICATION NUMBER: 60/080165
;; PRIOR FILING DATE: 1998-03-31
;; PRIOR APPLICATION NUMBER: 60/080194
;; PRIOR FILING DATE: 1998-03-31
;; PRIOR APPLICATION NUMBER: 60/080327
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/080328
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/080333
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/080334
;; PRIOR FILING DATE: 1998-04-01
;; PRIOR APPLICATION NUMBER: 60/081070
;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081049
;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081071
;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081195
;; PRIOR FILING DATE: 1998-04-08
;; PRIOR APPLICATION NUMBER: 60/081203
;; PRIOR FILING DATE: 1998-04-09
;; PRIOR APPLICATION NUMBER: 60/081229
;; PRIOR FILING DATE: 1998-04-09
;; PRIOR APPLICATION NUMBER: 60/081955
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/081957
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/081819
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/081952
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/082569
;; PRIOR FILING DATE: 1998-04-21
;; PRIOR APPLICATION NUMBER: 60/082704
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/082804
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/082700
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/082797
;; PRIOR FILING DATE: 1998-04-22
;; PRIOR APPLICATION NUMBER: 60/082796
;; PRIOR FILING DATE: 1998-04-23
;; PRIOR APPLICATION NUMBER: 60/083336
;; PRIOR FILING DATE: 1998-04-27
;; PRIOR APPLICATION NUMBER: 60/083322
;; PRIOR FILING DATE: 1998-04-28
;; PRIOR APPLICATION NUMBER: 60/083392
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083495
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083496
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083499
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083545

;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083554
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083558
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083559
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083500
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083742
;; PRIOR FILING DATE: 1998-04-30
;; PRIOR APPLICATION NUMBER: 60/084366
;; PRIOR FILING DATE: 1998-05-05
;; PRIOR APPLICATION NUMBER: 60/084414
;; PRIOR FILING DATE: 1998-05-06
;; PRIOR APPLICATION NUMBER: 60/084441
;; PRIOR FILING DATE: 1998-05-06
;; PRIOR APPLICATION NUMBER: 60/084637
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084639
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084640
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084598
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084600
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084627
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084643
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/085339
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085338
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085323
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085582
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085700
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085689
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085579
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085580
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085573
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085704
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085697

Query Match 1.6%; Score 7; DB 10; Length 472;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 KKEKPEE 285
DB 183 KKEKPEE 189

RESULT 35
US-09-978-192A-245
; Sequence 245; Application US/09978192A
; Patent No. US20020177553A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Deenoyers, Luc
; APPLICANT: Baton, Dan
; APPLICANT: Ferrara, Napoleon

APPLICANT: Filvaroff, Ellen
APPLICANT: Bong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerltsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Guiney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James J.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630P1C9
CURRENT APPLICATION NUMBER: US/09/978,192A
PRIOR FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079234
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079669
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923

PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29

PRIOR APPLICATION NUMBER: 60/083500
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083742
 PRIOR FILING DATE: 1998-04-30
 PRIOR APPLICATION NUMBER: 60/084366
 PRIOR FILING DATE: 1998-05-05
 PRIOR APPLICATION NUMBER: 60/084414
 PRIOR FILING DATE: 1998-05-06
 PRIOR APPLICATION NUMBER: 60/084441
 PRIOR FILING DATE: 1998-05-06
 PRIOR APPLICATION NUMBER: 60/084637
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/084639
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/084640
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/084598
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/084600
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/084627
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/084643
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/085339
 PRIOR FILING DATE: 1998-05-13
 PRIOR APPLICATION NUMBER: 60/085338
 PRIOR FILING DATE: 1998-05-13
 PRIOR APPLICATION NUMBER: 60/085323
 PRIOR FILING DATE: 1998-05-13
 PRIOR APPLICATION NUMBER: 60/085582
 PRIOR FILING DATE: 1998-05-15
 PRIOR APPLICATION NUMBER: 60/085700
 PRIOR FILING DATE: 1998-05-15
 PRIOR APPLICATION NUMBER: 60/085689
 PRIOR FILING DATE: 1998-05-15
 PRIOR APPLICATION NUMBER: 60/085579
 PRIOR FILING DATE: 1998-05-15
 PRIOR APPLICATION NUMBER: 60/085580
 PRIOR FILING DATE: 1998-05-15
 PRIOR APPLICATION NUMBER: 60/085573
 PRIOR FILING DATE: 1998-05-15
 PRIOR APPLICATION NUMBER: 60/085704
 PRIOR FILING DATE: 1998-05-15
 PRIOR APPLICATION NUMBER: 60/085697

Query Match 1.6%; Score 7; DB 10; Length 472;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 279 KKEKPEE 285
 Db 183 KKEKPEE 189

RESULT 36
 US-09-999-832A-245
 Sequence 245. Application US/09999832A

Publication No. US20020192706A1
 GENERAL INFORMATION:

APPLICANT: Ashkenazi, Avi
 APPLICANT: Baker Kevin P.
 APPLICANT: Botstein, David
 APPLICANT: Desnoyers, Luc
 APPLICANT: Eaton, Dan
 APPLICANT: Ferrara, Napoleon
 APPLICANT: Filvaroff, Ellen
 APPLICANT: Fong, Sherman
 APPLICANT: Gao, Wei-Qiang
 APPLICANT: Gerber, Hanspeter
 APPLICANT: Gerritsen, Mary E.
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, J. Christopher
 APPLICANT: Gurney, Austin L.
 APPLICANT: Hillan, Kenneth J.
 APPLICANT: Kijavlin, Ivar J.
 APPLICANT: Kuo, Sophia S.
 APPLICANT: Napier, Mary A.
 APPLICANT: Pan, James
 APPLICANT: Paoni, Nicholas F.
 APPLICANT: Roy, Margaret Ann
 APPLICANT: Shelton, David L.
 APPLICANT: Stewart, Timothy A.
 APPLICANT: Thomas, Daniel
 APPLICANT: Williams, P. Mickey
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 TITLE OF INVENTION: Acids Encoding the Same
 FILE REFERENCE: P2630P1C63
 CURRENT APPLICATION NUMBER: US/09/999, 832A
 PRIOR FILING DATE: 2001-10-24
 PRIOR APPLICATION NUMBER: 09/918585
 PRIOR FILING DATE: 2001-07-30
 PRIOR APPLICATION NUMBER: 60/062250
 PRIOR FILING DATE: 1997-10-17
 PRIOR APPLICATION NUMBER: 60/064249
 PRIOR FILING DATE: 1997-11-03
 PRIOR APPLICATION NUMBER: 60/065311
 PRIOR FILING DATE: 1997-11-13
 PRIOR APPLICATION NUMBER: 60/066364
 PRIOR FILING DATE: 1997-11-21
 PRIOR APPLICATION NUMBER: 60/077450
 PRIOR FILING DATE: 1998-03-10
 PRIOR APPLICATION NUMBER: 60/077632
 PRIOR FILING DATE: 1998-03-11
 PRIOR APPLICATION NUMBER: 60/077641
 PRIOR FILING DATE: 1998-03-11
 PRIOR APPLICATION NUMBER: 60/077649
 PRIOR FILING DATE: 1998-03-11
 PRIOR APPLICATION NUMBER: 60/077791
 PRIOR FILING DATE: 1998-03-12
 PRIOR APPLICATION NUMBER: 60/078004
 PRIOR FILING DATE: 1998-03-13
 PRIOR APPLICATION NUMBER: 60/078886
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/078936
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/078910
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/078939
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/079294
 PRIOR FILING DATE: 1998-03-25
 PRIOR APPLICATION NUMBER: 60/079656
 PRIOR FILING DATE: 1998-03-26
 PRIOR APPLICATION NUMBER: 60/079664
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/079689
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/079663
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/079728
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/079786
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/079920
 PRIOR FILING DATE: 1998-03-30
 PRIOR APPLICATION NUMBER: 60/079923
 PRIOR FILING DATE: 1998-03-30
 PRIOR APPLICATION NUMBER: 60/080105
 PRIOR FILING DATE: 1998-03-31
 PRIOR APPLICATION NUMBER: 60/080107
 PRIOR FILING DATE: 1998-03-31
 PRIOR APPLICATION NUMBER: 60/080165
 PRIOR FILING DATE: 1998-03-31

PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414

PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084598
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085689
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085697

Query Match 1.6%; Score 7; DB 10; Length 472;
Best Local Similarity 100.0%; Pred. No. 3e+02; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 279 KKEPPE 285
Db 183 KKEPPE 189

RESULT 37
US-09-978-189-245
Sequence 245, Application US/0978189
Publication No. US20030004102A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi
APPLICANT: Baker Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gueney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James;

APPLICANT: Paoni, Nicholas F.
 APPLICANT: Roy, Margaret Ann
 APPLICANT: Shelton, David L.
 APPLICANT: Stewart, Timothy A.
 APPLICANT: Thomas, Daniel
 APPLICANT: Williams, P. Mickey
 APPLICANT: Wood, William I.
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 FILE REFERENCE: P2630P1C7
 CURRENT APPLICATION NUMBER: US/09/978,189
 PRIOR FILING DATE: 2001-10-15
 PRIOR APPLICATION NUMBER: 09/918585
 PRIOR FILING DATE: 2001-07-30
 PRIOR APPLICATION NUMBER: 60/062250
 PRIOR FILING DATE: 1997-10-17
 PRIOR APPLICATION NUMBER: 60/064249
 PRIOR FILING DATE: 1997-11-03
 PRIOR APPLICATION NUMBER: 60/065311
 PRIOR FILING DATE: 1997-11-13
 PRIOR APPLICATION NUMBER: 60/066364
 PRIOR FILING DATE: 1997-11-21
 PRIOR APPLICATION NUMBER: 60/077450
 PRIOR FILING DATE: 1998-03-10
 PRIOR APPLICATION NUMBER: 60/077632
 PRIOR FILING DATE: 1998-03-11
 PRIOR APPLICATION NUMBER: 60/077641
 PRIOR FILING DATE: 1998-03-11
 PRIOR APPLICATION NUMBER: 60/077649
 PRIOR FILING DATE: 1998-03-11
 PRIOR APPLICATION NUMBER: 60/077791
 PRIOR FILING DATE: 1998-03-12
 PRIOR APPLICATION NUMBER: 60/078004
 PRIOR FILING DATE: 1998-03-13
 PRIOR APPLICATION NUMBER: 60/078886
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/078936
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/078910
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/078939
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/079294
 PRIOR FILING DATE: 1998-03-25
 PRIOR APPLICATION NUMBER: 60/079656
 PRIOR FILING DATE: 1998-03-26
 PRIOR APPLICATION NUMBER: 60/079664
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/079689
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/079663
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/079728
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/079786
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/079920
 PRIOR FILING DATE: 1998-03-30
 PRIOR APPLICATION NUMBER: 60/079923
 PRIOR FILING DATE: 1998-03-30
 PRIOR APPLICATION NUMBER: 60/080105
 PRIOR FILING DATE: 1998-03-31
 PRIOR APPLICATION NUMBER: 60/080107
 PRIOR FILING DATE: 1998-03-31
 PRIOR APPLICATION NUMBER: 60/080165
 PRIOR FILING DATE: 1998-03-31
 PRIOR APPLICATION NUMBER: 60/080194
 PRIOR FILING DATE: 1998-03-31
 PRIOR APPLICATION NUMBER: 60/080327
 PRIOR FILING DATE: 1998-04-01
 PRIOR APPLICATION NUMBER: 60/080328
 PRIOR FILING DATE: 1998-04-01
 PRIOR APPLICATION NUMBER: 60/080333

PRIOR FILING DATE: 1998-04-01
 PRIOR APPLICATION NUMBER: 60/080334
 PRIOR FILING DATE: 1998-04-01
 PRIOR APPLICATION NUMBER: 60/081070
 PRIOR FILING DATE: 1998-04-08
 PRIOR APPLICATION NUMBER: 60/081049
 PRIOR FILING DATE: 1998-04-08
 PRIOR APPLICATION NUMBER: 60/081071
 PRIOR FILING DATE: 1998-04-08
 PRIOR APPLICATION NUMBER: 60/081195
 PRIOR FILING DATE: 1998-04-08
 PRIOR APPLICATION NUMBER: 60/081203
 PRIOR FILING DATE: 1998-04-09
 PRIOR APPLICATION NUMBER: 60/081229
 PRIOR FILING DATE: 1998-04-09
 PRIOR APPLICATION NUMBER: 60/081955
 PRIOR FILING DATE: 1998-04-15
 PRIOR APPLICATION NUMBER: 60/081817
 PRIOR FILING DATE: 1998-04-15
 PRIOR APPLICATION NUMBER: 60/081819
 PRIOR FILING DATE: 1998-04-15
 PRIOR APPLICATION NUMBER: 60/081952
 PRIOR FILING DATE: 1998-04-15
 PRIOR APPLICATION NUMBER: 60/081838
 PRIOR FILING DATE: 1998-04-15
 PRIOR APPLICATION NUMBER: 60/082568
 PRIOR FILING DATE: 1998-04-21
 PRIOR APPLICATION NUMBER: 60/082569
 PRIOR FILING DATE: 1998-04-21
 PRIOR APPLICATION NUMBER: 60/082704
 PRIOR FILING DATE: 1998-04-22
 PRIOR APPLICATION NUMBER: 60/082804
 PRIOR FILING DATE: 1998-04-22
 PRIOR APPLICATION NUMBER: 60/082700
 PRIOR FILING DATE: 1998-04-22
 PRIOR APPLICATION NUMBER: 60/082797
 PRIOR FILING DATE: 1998-04-22
 PRIOR APPLICATION NUMBER: 60/082796
 PRIOR FILING DATE: 1998-04-23
 PRIOR APPLICATION NUMBER: 60/083336
 PRIOR FILING DATE: 1998-04-27
 PRIOR APPLICATION NUMBER: 60/083322
 PRIOR FILING DATE: 1998-04-28
 PRIOR APPLICATION NUMBER: 60/083392
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083495
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083496
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083499
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083545
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083554
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083558
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083559
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083500
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/083742
 PRIOR FILING DATE: 1998-04-30
 PRIOR APPLICATION NUMBER: 60/084366
 PRIOR FILING DATE: 1998-05-05
 PRIOR APPLICATION NUMBER: 60/084414
 PRIOR FILING DATE: 1998-05-06
 PRIOR APPLICATION NUMBER: 60/084441
 PRIOR FILING DATE: 1998-05-06
 PRIOR APPLICATION NUMBER: 60/084637
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/084639
 PRIOR FILING DATE: 1998-05-07

```

; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084598
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085689
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697
```

```

Query Match          1.6%; Score 7; DB 11; Length 472;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      279 KKEKPEE 285
Db      183 KKEKPEE 189
```

```

RESULT 38
US-09-978-608A-245
; Sequence 245, Application US/09978608A
; Publication No. US20030045462A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary B.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James;
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
```

```

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2630PIC22
; CURRENT APPLICATION NUMBER: US/09/978,608A
; CURRENT FILING DATE: 2001-10-16
; NUMBER OF SEQ ID NOS: 624
; Prior Application removed - See File Wrapper or Palm
; SEQ ID NO 245
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-978-608A-245
```

```

Query Match          1.6%; Score 7; DB 11; Length 472;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      279 KKEKPEE 285
Db      183 KKEKPEE 189
```

```

RESULT 39
US-09-978-585A-245
; Sequence 245, Application US/09978585A
; Publication No. US20030049633A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary B.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James;
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2630PIC15
; CURRENT APPLICATION NUMBER: US/09/978,585A
; CURRENT FILING DATE: 2001-10-16
; NUMBER OF SEQ ID NOS: 624
; Prior Application removed - See File Wrapper or Palm
; SEQ ID NO 245
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-978-585A-245
```

```

Query Match          1.6%; Score 7; DB 11; Length 472;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      279 KKEKPEE 285
Db      183 KKEKPEE 189
```


Db 183 KKKKPEE 189

RESULT 40
US-09-978-191A-245
Sequence 245, Application US/09978191A
Publication No. US20030050239A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi
APPLICANT: Baker Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertlisen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James:
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secretd and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2630P1C4
CURRENT APPLICATION NUMBER: US/09/978,191A
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/06364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/07450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656

PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29

PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084598
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085689
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085697

Query Match 1.6%; Score 7; DB 11; Length 472;
Best Local Similarity 100.0%; Pred.No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 KKEKPEE 285
|||
Db 183 KKEKPEE 189

Search completed: August 8, 2003, 10:01:52
Job time : 53 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2003, 09:48:57 ; Search time 29 Seconds
(without alignments)
630.266 Million cell updates/sec

Title: US-10-069-062-7

Perfect score: 432
Sequence: 1 MSKFSAPGKAFIAGYIVL.....DIEQTEGVLEKPEYDYLGL 432

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8.1	9.398	4	US-09-328-352-6136	Sequence 6136, App
2	7.1	6.123	4	US-09-252-991A-23118	Sequence 23118, A
3	7.1	6.172	4	US-09-328-352-6827	Sequence 6827, App
4	7.1	6.176	3	US-09-415-523-10	Sequence 10, Appl
5	7.1	6.220	1	US-08-033-797-2	Sequence 2, Appl
6	7.1	6.220	1	US-08-472-265-2	Sequence 2, Appl
7	7.1	6.220	1	US-08-472-265-2	Sequence 2, Appl
8	7.1	6.258	3	US-09-068-140A-6	Sequence 6, Appl
9	7.1	6.284	4	US-09-056-556-162	Sequence 162, App
10	7.1	6.284	4	US-09-072-596-157	Sequence 157, App
11	7.1	6.291	3	US-08-446-137B-4	Sequence 4, Appl
12	7.1	6.302	4	US-09-328-352-4508	Sequence 4508, App
13	7.1	6.374	3	US-09-091-405-2	Sequence 2, Appl
14	7.1	6.405	4	US-09-413-574-2	Sequence 2, Appl
15	7.1	6.416	4	US-09-252-991A-17015	Sequence 17015, A
16	7.1	6.426	4	US-09-134-001C-3346	Sequence 3346, App
17	7.1	6.472	3	US-09-088-425-1	Sequence 1, Appl
18	7.1	6.472	4	US-09-639-378A-1	Sequence 140, App
19	7.1	6.488	4	US-09-477-135A-140	Sequence 2, Appl
20	7.1	6.508	3	US-08-991-677-2	Sequence 2, Appl
21	7.1	6.567	2	US-08-504-459-2	Sequence 2, Appl
22	7.1	6.592	4	US-09-423-468A-15	Sequence 15, Appl
23	7.1	6.592	6	5200183-18	Patent No. 5200183
24	7.1	6.594	4	US-09-252-991A-19275	Sequence 19275, A
25	7.1	6.631	4	US-09-147-119-7	Sequence 7, Appl
26	7.1	6.675	4	US-09-252-991A-23185	Sequence 23185, A
27	7.1	6.1027	3	US-08-446-137B-2	Sequence 2, Appl

28	7	1.6	1264	4	US-09-252-991A-30264	Sequence 30264, A
29	7	1.6	2470	4	US-08-265-967C-2	Sequence 2, Appl
30	7	1.6	2470	4	US-08-305-790B-3	Sequence 3, Appl
31	6	1.4	9	2	US-08-753-781-30	Sequence 30, Appl
32	6	1.4	11	2	US-08-459-568-6	Sequence 6, Appl
33	6	1.4	11	2	US-08-399-411-6	Sequence 6, Appl
34	6	1.4	11	3	US-08-516-859A-6	Sequence 6, Appl
35	6	1.4	11	3	US-09-586-472-6	Sequence 6, Appl
36	6	1.4	11	4	US-09-528-706-6	Sequence 6, Appl
37	6	1.4	12	3	US-08-464-841A-1	Sequence 1, Appl
38	6	1.4	15	4	US-09-009-953-241	Sequence 241, App
39	6	1.4	15	4	US-09-311-784A-427	Sequence 427, App
40	6	1.4	16	3	US-08-464-841A-3	Sequence 3, Appl
41	6	1.4	18	4	US-09-263-312-4	Sequence 4, Appl
42	6	1.4	20	2	US-08-568-310D-7	Sequence 7, Appl
43	6	1.4	25	2	US-09-270-455-7	Sequence 7, Appl
44	6	1.4	20	4	US-08-473-475A-20	Sequence 20, Appl
45	6	1.4	31	5	PCT-US93-11703-63	Sequence 63, Appl

ALIGNMENTS

RESULT 1
US-09-328-352-6136
; Sequence 6136, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6136
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6136

Query Match 1.9%; Score 8; DB 4; Length 398;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 SAGIWSV 157
DB 313 SAGIWSV 320

RESULT 2
US-09-252-991A-23118
; Sequence 23118, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23118
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23118

Query Match 1.6%; Score 7; DB 4; Length 123;

Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 AGLVSVV 157
|||||

DB 53 AGLVSVV 59

RESULT 3

US-09-328-352-6827
; Sequence 6827, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6827
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6827

Query Match 1.6%; Score 7; DB 4; Length 172;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 KGGSETP 269
|||||

DB 113 KGGSETP 119

RESULT 4

US-09-415-522-10
; Sequence 10, Application US/09415522A
; Patent No. 6291660
; GENERAL INFORMATION:
; APPLICANT: Gaffney, Thomas
; APPLICANT: Wendland, Juergen
; APPLICANT: Philippesen, Peter
; TITLE OF INVENTION: No. 6291660e1 Fungal Genes Required For No. 6291660e1 Growth And
; FILE REFERENCE: CGC2045
; CURRENT APPLICATION NUMBER: US/09/415,522A
; CURRENT FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Ashbya gossypii
US-09-415-522-10

Query Match 1.6%; Score 7; DB 3; Length 176;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 328 PLTVAIK 334
|||||

DB 55 PLTVAIK 61

RESULT 5

US-08-033-797-2
; Sequence 2, Application US/08033797
; Patent No. 5583210
; GENERAL INFORMATION:
; APPLICANT: NEILL, John
; APPLICANT: PIERCE, Dorothy A.
; APPLICANT: CIGAN, Andrew M.

TITLE OF INVENTION: Methods and Compositions for Controlling
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/033,797
FILING DATE: 18-MAR-1993
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 33229/168/PIHI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 220 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-033-797-2

Query Match 1.6%; Score 7; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 APGKAPL 13
|||||

DB 210 APGKAPL 216

RESULT 6

US-08-472-265-2
; Sequence 2, Application US/08472265
; Patent No. 5728817
; GENERAL INFORMATION:
; APPLICANT: NEILL, John
; APPLICANT: PIERCE, Dorothy A.
; APPLICANT: CIGAN, Andrew M.
; TITLE OF INVENTION: Methods and Compositions for Controlling
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,265
FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/033,797
FILING DATE: 18-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 33229/335/PIHI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 220 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-472-265-2

Query Match 1.6%; Score 7; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 AFGKAPL 13
|||||
DB 210 AFGKAPL 216

RESULT 7
US-08-472-263-2
Sequence 2, Application US/08472263
Patent No. 5760190
GENERAL INFORMATION:
APPLICANT: NEILL, John
APPLICANT: PIERCE, Dorothy A.
APPLICANT: CIGAN, Andrew M.
TITLE OF INVENTION: Methods and Compositions for Controlling
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,263
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/033,797
FILING DATE: 18-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 33229/334/PIHI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 220 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-472-263-2

Query Match 1.6%; Score 7; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 AFGKAPL 13
|||||

DB 210 AFGKAPL 216

RESULT 8
US-09-068-140A-6
Sequence 6, Application US/09068140A
Patent No. 6281409
GENERAL INFORMATION:
APPLICANT: Mary Rose Woodhead, Mark Andrew Taylor
APPLICANT: and Rex Michael Brennan
TITLE OF INVENTION: Blackcurrant Promoters and Genes
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/068,140A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP96/04807
FILING DATE: No. 6281409ember 4, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Dimmer, Dara L.
REGISTRATION NUMBER: 33,680
REFERENCE/DOCKET NUMBER: C70237
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5017
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 258 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Ribes nigrum
STRAIN: Ben Alder
US-09-068-140A-6

Query Match 1.6%; Score 7; DB 3; Length 258;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 DHSVEPL 329
|||||
DB 150 DHSVEPL 156

RESULT 9
US-09-056-556-162
Sequence 162, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yaelir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241

TREATM

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: SEED and BERRY LLP
;; STREET: 6300 Columbia Center, 701 Fifth Avenue
;; CITY: Seattle
;; STATE: Washington
;; COUNTRY: USA
;; ZIP: 98104-7092
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/056,556
;; FILING DATE: 07-Apr-1998
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Maki, David J.
;; REGISTRATION NUMBER: 31,392
;; REFERENCE/DOCKET NUMBER: 210121.457
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206) 622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ ID NO: 162:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 284 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; US-09-056-556-162

Query Match 1.6%; Score 7; DB 4; Length 284;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 380 YDAIAVL 386
DB 17 YDAIAVL 23

RESULT 10
US-09-072-596-157
; Sequence 157, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Nero, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

;; NAME: Maki, David J.
;; REGISTRATION NUMBER: 31,392
;; REFERENCE/DOCKET NUMBER: 210121.417C9
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206) 622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ ID NO: 157:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 284 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; US-09-072-596-157

Query Match 1.6%; Score 7; DB 4; Length 284;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 380 YDAIAVL 386
DB 17 YDAIAVL 23

RESULT 11
US-08-446-137B-4
; Sequence 4, Application US/08446137B
; Patent No. 6162903
; GENERAL INFORMATION:
; APPLICANT: Townen, Angus R.
; APPLICANT: Atkinson, Anthony
; APPLICANT: Murphy, Jonathan P.
; APPLICANT: Laurence, Oliver S.
; APPLICANT: Dugleby, Clive J.
; TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED
; FROM L PROTEIN AND THEIR USES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,137B
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: McMaisters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.406
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULAR TYPE: protein
; US-08-446-137B-4

Query Match 1.6%; Score 7; DB 3; Length 291;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 279 KKEKPEE 285
|||||||

Db 71 KKEKPEE 77

RESULT 12

US-09-328-352-4508
; Sequence 4508, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; EARLIER FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4508
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4508

Query Match 1.6%; Score 7; DB 4; Length 302;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 QUNSANL 298
|||||
Db 164 QUNSANL 170

RESULT 13

US-09-091-405-2
; Sequence 2, Application US/09091405A
; Patent No. 6207144
; GENERAL INFORMATION:
; APPLICANT: KORTH, Reinhard
; APPLICANT: BAIER, Michael
; APPLICANT: BANNERT, No. 6207144bert
; APPLICANT: METZNER, Karin
; APPLICANT: WERNER, Albrecht
; APPLICANT: LANG, Kurt
; TITLE OF INVENTION: POLYPEPTIDES WITH INTERLEUKIN 16 ACTIVITY, PROCESS FOR
; FILE REFERENCE: 8341-8041
; CURRENT APPLICATION NUMBER: US/09/091,405A
; EARLIER FILING DATE: 1998-08-27
; EARLIER APPLICATION NUMBER: PCT/EP96/05662
; EARLIER FILING DATE: 1996-12-17
; EARLIER APPLICATION NUMBER: DE/196 13 886.8
; EARLIER FILING DATE: 1996-04-06
; EARLIER APPLICATION NUMBER: DE/196 13 866.3
; EARLIER FILING DATE: 1996-04-06
; EARLIER APPLICATION NUMBER: DE/196 03 492.2
; EARLIER FILING DATE: 1996-01-31
; EARLIER APPLICATION NUMBER: DE/195 48 295.6
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: IL-16
US-09-091-405-2

Query Match 1.6%; Score 7; DB 3; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ELKKIE 239
|||||
Db 131 ELKKIE 137

RESULT 14

US-09-413-574-2
; Sequence 2, Application US/09413574
; Patent No. 6235972
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; APPLICANT: Taghian, Laura
; TITLE OF INVENTION: Maize Rad23 Genes and Uses Thereof
; FILE REFERENCE: 0964
; CURRENT APPLICATION NUMBER: US/09/413,574
; EARLIER FILING DATE: 1999-10-06
; EARLIER APPLICATION NUMBER: 60/109,728
; EARLIER FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Zea mays
US-09-413-574-2

Query Match 1.6%; Score 7; DB 3; Length 405;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 GVPBGAG 378
|||||
Db 266 GVPBGAG 272

RESULT 15

US-09-252-991A-17015
; Sequence 17015, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17015
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17015

Query Match 1.6%; Score 7; DB 4; Length 416;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AFSAPGK 10
|||||
Db 358 AFSAPGK 364

RESULT 16

US-09-134-001C-3346
; Sequence 3346, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3346
LENGTH: 426
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3346

Query Match 1.6%; Score 7; DB 4; Length 426;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 KKEKPEE 49
DB 168 GTSLSKES 174

RESULT 17
US-09-088-425-1
Sequence 1, Application US/09088425
Patent No. 6171843
GENERAL INFORMATION:
APPLICANT: BANDMAN, OLGA
APPLICANT: LAL, PREETI
APPLICANT: CORLEY, NEIL C.
APPLICANT: PATTERSON, CHANDRA
APPLICANT: BAUGHN, MARIAN R.
TITLE OF INVENTION: HUMAN ISOMERASE-LIKE PROTEINS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/088,425
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Cetrone, Michael C
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0529 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRAINON01
CLONE: 2291164
US-09-088-425-1

Query Match 1.6%; Score 7; DB 3; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 279 KKEKPEE 285
DB 183 KKEKPEE 189

RESULT 18
US-09-639-378A-1
Sequence 1, Application US/09639378A
Patent No. 6524838
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil, C.
APPLICANT: Patterson, Chandra
APPLICANT: Baughn, Marian R.
TITLE OF INVENTION: HUMAN PEPTIDYL-PROLYL ISOMERASES
FILE REFERENCE: PF-0529-1 DIV
CURRENT APPLICATION NUMBER: US/09/639,378A
CURRENT FILING DATE: 2000-08-15
PRIOR APPLICATION NUMBER: US 09/088,425
PRIOR FILING DATE: 1998-06-01
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PERL Program
SEQ ID NO 1
LENGTH: 472
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No. 6524838 2291164CD1
US-09-639-378A-1

Query Match 1.6%; Score 7; DB 4; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 KKEKPEE 285
DB 183 KKEKPEE 189

RESULT 19
US-09-477-135A-140
Sequence 140, Application US/09477135A
Patent No. 6572865
GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
FILE REFERENCE: 5288
CURRENT APPLICATION NUMBER: US/09/477,135A
CURRENT FILING DATE: 2000-01-03
PRIOR APPLICATION NUMBER: 08990823
PRIOR FILING DATE: 1997-12-15
PRIOR APPLICATION NUMBER: US 96/10375
PRIOR FILING DATE: 1996-06-14
PRIOR APPLICATION NUMBER: 60/000,254
PRIOR FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 169
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 140
LENGTH: 488
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-477-135A-140

Query Match 1.6%; Score 7; DB 4; Length 488;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 156 VVATSL 162

Db 432 VVATSL 438

RESULT 20
US-08-991-677-2
Sequence 2, Application US/08991677A
Patent No. 6252135
GENERAL INFORMATION:
APPLICANT: Chiang, Vincent L
APPLICANT: Carriway, Daniel T
APPLICANT: Smeltzer, Richard H
TITLE OF INVENTION: Production of Syringyl Lignin in Gymnosperms
FILE REFERENCE: 50617
CURRENT APPLICATION NUMBER: US/08/991,677A
CURRENT FILING DATE: 1997-12-16
EARLIER APPLICATION NUMBER: US 60/033,381
EARLIER FILING DATE: 1996-12-16
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 508
TYPE: prt
ORGANISM: Liquidambar styraciflua
US-08-991-677-2

Query Match 1.6%; Score 7; DB 3; Length 508;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 IFIVLAY 95
Db 11 IFIVLAY 17

RESULT 21
US-08-504-459-2
Sequence 2, Application US/08504459
Patent No. 5922563
GENERAL INFORMATION:
APPLICANT: Alderete, John F.
TITLE OF INVENTION: Adhesin Genes and Proteins Involved in
TITLE OF INVENTION: Trichomonas Vaginalis Cytoadherence
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: US
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/504,459
FILING DATE: Concurrently herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mark B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: UTSK:273/WIM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
TELEX: N/A
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 567 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-504-459-2

Query Match 1.6%; Score 7; DB 2; Length 567;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 PKGTSLK 47
Db 543 PKGTSLK 549

RESULT 22
US-09-423-468A-15
Sequence 15, Application US/09423468A
Patent No. 6469149
GENERAL INFORMATION:
APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnol
APPLICANT: Guy Jerome Corneel Bauw
APPLICANT: Mark William Davey
APPLICANT: Jens Oostergaard
APPLICANT: Marc Charles Ernest Van Montagu
TITLE OF INVENTION: PRODUCTION OF ASCORBIC ACID IN PLANTS
FILE REFERENCE: DECELS.001APC
CURRENT APPLICATION NUMBER: US/09/423,468A
CURRENT FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: NL 1006000
PRIOR FILING DATE: 1997-05-07
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15
LENGTH: 592
TYPE: prt
ORGANISM: Arabidopsis thaliana
US-09-423-468A-15

Query Match 1.6%; Score 7; DB 4; Length 592;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 ELKKLIE 239
Db 476 ELKKLIE 482

RESULT 23
5200183-18
Patent No. 5200183
APPLICANT: TANG, JORDAN J.N.;WANG, CHI-SUN
TITLE OF INVENTION: RECOMBINANT BILE SALT ACTIVATED LIPASES
NUMBER OF SEQUENCES: 22
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,426
FILING DATE: 12-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 504,635
FILING DATE: 04-APR-1990
APPLICATION NUMBER: 122,410
FILING DATE: 19-NOV-1987
SEQ ID NO:18
LENGTH: 592
5200183-18

Query Match 1.6%; Score 7; DB 6; Length 592;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 374 VPGAGGY 380
Db 511 VPGAGGY 517

RESULT 24

US-09-252-991A-19275
; Sequence 19275, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19275
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19275

Query Match 1.6%; Score 7; DB 4; Length 594;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 EKKEDY 429
|||||
DB 503 EKKEDY 509

RESULT 25
US-09-147-119-7
; Sequence 7, Application US/09147119
; Patent No. 6338844

; GENERAL INFORMATION:
; APPLICANT: KURTH, Ruth
; APPLICANT: BAUER, Michael
; APPLICANT: BANNER, No. 6338844bert
; APPLICANT: METZNER, Karin
; APPLICANT: WERNER, Albrecht
; TITLE OF INVENTION: GENOMIC NUCLEIC ACIDS, CDNA AND mRNA WHICH
; TITLE OF INVENTION: CODE FOR POLYPEPTIDES WITH IL-16 ACTIVITY, PROCESS FOR THE
; NUMBER OF INVENTION: PRODUCTION THEREOF AND THEIR USE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikaido, Martelstein, Murray & Oram LLP
; STREET: 655 Fifteenth St., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/147,119
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP97/01753
; FILING DATE: 10-APR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 14 099.4
; FILING DATE: 10-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Wong, King L.
; REGISTRATION NUMBER: 37,500
; REFERENCE/DOCKET NUMBER: P1614-8068
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-5000
; TELEFAX: (202) 638-4810

; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 631 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-147-119-7

Query Match 1.6%; Score 7; DB 4; Length 631;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ELKKLIE 239
|||||
DB 388 ELKKLIE 394

RESULT 26
US-09-252-991A-23185
; Sequence 23185, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23185
; LENGTH: 675
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23185

Query Match 1.6%; Score 7; DB 4; Length 675;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 GGVPGA 377
|||||
DB 532 GGVPGA 538

RESULT 27
US-08-446-137B-2
; Sequence 2, Application US/08446137B
; Patent No. 6162903
; GENERAL INFORMATION:
; APPLICANT: Townen, Angus R.
; APPLICANT: Atkinson, Anthony
; APPLICANT: Murphy, Jonathan P.
; APPLICANT: Laurence, Oliver S.
; APPLICANT: Dugdaley, Clive J.
; TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED
; TITLE OF INVENTION: FROM L PROTEIN AND THEIR USES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,137B
FILING DATE: 22-MAY-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: McMaisters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 100084,406
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1027 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-446-137B-2

Query Match 1.6%; Score 7; DB 3; Length 1027;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 KKEKPEE 285
|||||
DB 353 KKEKPEE 359

RESULT 28
US-09-252-991A-30264
Sequence 30264, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 30264
LENGTH: 1264
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30264

Query Match 1.6%; Score 7; DB 4; Length 1264;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 VATSLLS 163
|||||
DB 903 VATSLLS 909

RESULT 29
US-08-265-967C-2
Sequence 2, Application US/08265967C
Patent No. 6476200
GENERAL INFORMATION:
APPLICANT: SABATINI, DAVID M.
APPLICANT: ERDJUMENT-BROMAGE, HEDIYE
APPLICANT: LUI, MARY
APPLICANT: TEMPEST, PAUL
APPLICANT: SNYDER, SOLOMON H.
TITLE OF INVENTION: MAMMALIAN PROTEINS THAT BIND TO FKBP12
TITLE OF INVENTION: IN A RAPAMYCIN-DEPENDENT FASHION
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:

ADDRESSEE: BANNER & ALLEGRETTI, LTD
STREET: 1001 G STREET, N.W., 11TH FLOOR
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20001-4597
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/265,967C
FILING DATE: 27-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KAGAN, SARAH A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107,46363
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 197430 BBNW UT
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2470 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
US-08-265-967C-2

Query Match 1.6%; Score 7; DB 4; Length 2470;
Best Local Similarity 100.0%; Pred. No. 7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 328 PLTVAIK 334
|||||
DB 1912 PLTVAIK 1918

RESULT 30
US-08-305-790B-3
Sequence 3, Application US/08305790B
Patent No. 6492106
GENERAL INFORMATION:
APPLICANT: SABATINI, DAVID M.
APPLICANT: ERDJUMENT-BROMAGE, HEDIYE
APPLICANT: LUI, MARY
APPLICANT: TEMPEST, PAUL
APPLICANT: SNYDER, SOLOMON H.
TITLE OF INVENTION: MAMMALIAN PROTEINS THAT BIND TO FKBP12
TITLE OF INVENTION: IN A RAPAMYCIN-DEPENDENT FASHION
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: BANNER & ALLEGRETTI, LTD
STREET: 1001 G STREET, N.W., 11TH FLOOR
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20001-4597
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/305,790B
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,967

FILING DATE: 27-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: KAGAN, SARAH A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107,47225
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 197430 BMB UT
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2470 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
US-08-305-790B-3

Query Match 1.6%; Score 7; DB 4; Length 2470;
Best Local Similarity 100.0%; Pred. No. 76+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 328 PLTVAIK 334
|||||
DB 1912 PLTVAIK 1918

RESULT 31
US-08-753-781-30
Sequence 30, Application US/08753781C
Patent No. 5951981
GENERAL INFORMATION:
APPLICANT: Markland Jr., Francis S.
APPLICANT: Bush, Larry R.
APPLICANT: Swenson, Stephen
APPLICANT: Flores Sanchez, Eladio
TITLE OF INVENTION: THROMBOLYTIC AGENTS WITH ANTITHROMBOTIC ACTIVITY
FILE REFERENCE: DITI 124
CURRENT APPLICATION NUMBER: US/08/753,781C
CURRENT FILING DATE: 1996-12-02
NUMBER OF SEQ ID NOS: 40
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 30
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (2)
OTHER INFORMATION: L-S-(3-aminopropyl) cysteine
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (9)
OTHER INFORMATION: AMIDATION
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
OTHER INFORMATION: peptide
US-08-753-781-30

Query Match 1.4%; Score 6; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 GDVKG 265
|||||
DB 3 GDVKG 8

RESULT 32
US-08-459-568-6
Sequence 6, Application US/08459568
Patent No. 5811304

GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
TITLE OF INVENTION: Zinc Finger Proteins
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSER: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,568
FILING DATE: 02-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-459-568-6

Query Match 1.4%; Score 6; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 EEKPD 428
|||||
DB 5 EEKPD 10

RESULT 33
US-08-399-411-6
Sequence 6, Application US/08399411
Patent No. 5831008
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
TITLE OF INVENTION: Zinc Finger Proteins
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSER: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/399,411
FILING DATE: 06-MAR-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-399-411-6

Query Match 1.4%; Score 6; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 EKKPED 428
|||||
DB 5 EKKPED 10

RESULT 34
US-08-516-859A-6
Sequence 6, Application US/08516859A
Patent No. 6069231
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSER: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/516,859A
FILING DATE: 18-AUG-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/292,683
FILING DATE: 18-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1276
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-516-859A-6

Query Match 1.4%; Score 6; DB 3; Length 11;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 423 EKKPED 428
|||||

DB 5 EKKPED 10

RESULT 35
US-09-586-472-6
Sequence 6, Application US/09586472
Patent No. 6323335
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSER: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/586,472
FILING DATE: 01-Jun-2000
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/528,706
FILING DATE: 17-MAR-2000
APPLICATION NUMBER: US 08/516,859
FILING DATE: 18-AUG-1995
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
APPLICATION NUMBER: US 08/292,683
FILING DATE: 18-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 4130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: amino acid
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-586-472-6

Query Match 1.4%; Score 6; DB 4; Length 11;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 423 EKKPED 428
|||||
DB 5 EKKPED 10

RESULT 36
US-09-528-706-6
Sequence 6, Application US/09528706
Patent No. 6468985
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSER: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/528,706
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/516,859
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/292,683
FILING DATE: 18-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1776
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-8949
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-09-528-706-6

Query Match 1.4%; Score 6; DB 4; Length 11;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 EKKED 428
DB 5 EKKED 10

RESULT 37
US-08-464-841A-1
Sequence 1, Application US/08464841A
Patent No. 6103880
GENERAL INFORMATION:
APPLICANT: BARRITULT, DENIS; COURT, JOSE;
TITLE OF INVENTION: HARP FAMILY GROWTH FACTORS, PREPARATION
TITLE OF INVENTION: METHODS THEREFOR AND USES THEREOF
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCIT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,841A
FILING DATE: 23 AUGUST 1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR94/00219
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN

REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 102.163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12
TYPE: Amino Acid
STRANDEDNESS: Unknown
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-464-841A-1

Query Match 1.4%; Score 6; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 KKEKE 284
DB 5 KKEKE 10

RESULT 38
US-09-009-953-241
Sequence 241, Application US/09009953
Patent No. 6413517
GENERAL INFORMATION:
APPLICANT: Sette, Alessandro
TITLE OF INVENTION: Identification of Broadly
Reactive DR Restricted Epitopes
NUMBER OF SEQUENCES: 274
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/009,953
FILING DATE: 21-Jan-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/036,713
FILING DATE: 23-JAN-1997
APPLICATION NUMBER: US 60/037,432
FILING DATE: 07-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen Lauver
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 018623-011520US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 241:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 241:
US-09-009-953-241
Query Match 1.4%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 67;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 SLKESR 50
|||||
Db 7 SLKESR 12

RESULT 39
US-09-311-784A-427
Sequence 427, Application US/09311784A
Patent No. 6534482
GENERAL INFORMATION:
APPLICANT: Fikes, John D.
APPLICANT: Hermanson, Gary G.
APPLICANT: Sette, Alessandro
APPLICANT: Ishioaka, Glenn Y.
APPLICANT: Livingston, Brian
APPLICANT: Chesnut, Robert W.
APPLICANT: Eptimmune, Inc.
TITLE OF INVENTION: Expression Vectors for Stimulating an
TITLE OF INVENTION: Immune Response and Methods of Using the Same
FILE REFERENCE: 39963-20022.01
CURRENT APPLICATION NUMBER: US/09/311,784A
PRIOR FILING DATE: 1999-05-13
PRIOR APPLICATION NUMBER: US 60/085,751
NUMBER OF SEQ ID NOS: 463
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 427
LENGTH: 15
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Pf SSP2 165 (peptide 35.0171)
US-09-311-784A-427

Query Match 1.4%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 SLKESR 50
|||||
Db 7 SLKESR 12

RESULT 40
US-08-464-841A-3
Sequence 3, Application US/08464841A
Patent No. 6103880
GENERAL INFORMATION:
APPLICANT: BARRITAU, DENIS; COURT, JOSE;
APPLICANT: LAAROUBI, KHALID
TITLE OF INVENTION: HARP FAMILY GROWTH FACTORS, PREPARATION
TITLE OF INVENTION: METHODS THEREFOR AND USES THEREOF
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,841A
FILING DATE: 23 AUGUST 1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR94/00219

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 102.163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 16
TYPE: Amino Acid
STRANDEDNESS: Unknown
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-464-841A-3

Query Match 1.4%; Score 6; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 KKEKPE 284
|||||
Db 9 KKEKPE 14

Search completed: August 8, 2003, 09:54:03
Job time : 31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2003, 09:34:31 ; Search time 100 Seconds
(without alignments)
114.787 Million cell updates/sec

Title: US-10-069-062-7

Perfect score: 2230
Sequence: 1 MSKAPSGKAFKLAGYIVL.....DLEQTEGVLEKPEDYICL 432

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	611.5	27.4	426	3 Q9UT88	Q9UT88 schizosacch
2	475	21.3	505	10 Q9C6T1	Q9C6T1 arabidopsis
3	445.5	20.0	503	10 Q944G1	Q944G1 hevea bras
4	247	11.1	369	16 Q8ETN1	Q8ETN1 oceanobact
5	228	10.2	314	17 Q973B5	Q973B5 sulfolobus
6	225	10.1	323	17 Q97UL6	Q97UL6 sulfolobus
7	209.5	9.4	368	2 Q9FD67	Q9FD67 enterococc
8	193	8.7	358	2 Q9FD77	Q9FD77 staphylococ
9	181.5	8.1	358	16 Q9FD72	Q9FD72 staphylococ
10	181	8.1	358	16 Q8NXW0	Q8NXW0 staphylococ
11	177.5	8.0	361	2 Q9FD62	Q9FD62 enterococc
12	173	7.8	358	2 Q9FD83	Q9FD83 staphylococ
13	173	7.8	358	16 Q99W20	Q99W20 staphylococ
14	170.5	7.6	359	16 Q8YAV1	Q8YAV1 listeria mo
15	165.5	7.4	360	16 Q92F01	Q92F01 listeria in
16	147	6.6	374	2 Q9KMG3	Q9KMG3 streptococ

17	141.5	6.3	294	2 Q93K31	Q93K31 lactobacill
18	129	5.8	317	16 Q51630	Q51630 borrelia bu
19	127	5.7	376	10 Q9STB1	Q9STB1 arabidopsis
20	122.5	5.5	3187	11 Q63714	Q63714 rattus norv
21	120	5.4	449	16 Q9PJ75	Q9PJ75 campylobact
22	118	5.3	640	16 Q9CKM9	Q9CKM9 pasteurella
23	117.5	5.3	1088	10 Q9FIM1	Q9FIM1 arabidopsis
24	117	5.2	335	16 Q97SH9	Q97SH9 streptococ
25	117	5.2	335	16 Q8DR49	Q8DR49 streptococ
26	116.5	5.2	352	5 Q8UBR3	Q8UBR3 encephalico
27	116	5.2	862	10 Q43191	Q43191 solanum tub
28	114	5.1	1342	2 Q9FKN5	Q9FKN5 arabidopsis
29	113.5	5.1	336	2 Q9FD52	Q9FD52 streptococ
30	113	5.1	340	2 Q8XK05	Q8XK05 streptococ
31	111.5	5.0	733	5 Q17102	Q17102 caenorhabdi
32	111	5.0	642	2 Q91423	Q91423 anabaena sp
33	111	5.0	1249	10 Q23289	Q23289 arabidopsis
34	110.5	5.0	332	16 Q8DU00	Q8DU00 streptococ
35	110.5	5.0	507	5 Q9VJ16	Q9VJ16 drosophila
36	110.5	5.0	600	17 Q29291	Q29291 archaeoglob
37	110.5	5.0	790	5 Q20949	Q20949 caenorhabdi
38	110.5	5.0	3853	5 Q81JW2	Q81JW2 plasmodium
39	110	4.9	664	17 Q58354	Q58354 pyrococcus
40	109.5	4.9	643	17 Q9UY87	Q9UY87 pyrococcus
41	109.5	4.9	1332	10 Q9S1G7	Q9S1G7 arabidopsis
42	109	4.9	587	17 Q8ZYP6	Q8ZYP6 pyrobaculum
43	108.5	4.9	1540	6 Q29433	Q29433 sus scrofa
44	107.5	4.8	739	5 Q9N1H1	Q9N1H1 plasmodium
45	107.5	4.8	743	5 Q815D2	Q815D2 plasmodium

ALIGNMENTS

RESULT 1	ID	Q9UT88	PRELIMINARY;	PRT;	426 AA.
AC	Q9UT88;	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)				
DE	Putative phosphoenolpyruvate kinase.				
GN	SPAC343.01C.				
OS	Schizosaccharomyces pombe (Fission Yeast).				
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;				
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;				
OX	NCBI_TaxID=4896;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=972h-;				
RA	Murphy L., Harris D., Wood V., Rajandream M.A., Barrell B.G.;				
RL	Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AL109739; CAB52264.1; -				
DR	GeneDB; SPombe; SPAC343.01C; -				
DR	InterPro; IPR005916; Pmev_kin_erg8.				
DR	TIGRFAms; TIGR01219; Pmev_kin_Erg8; 1.				
KW	kinase				
SC	SEQUENCE 426 AA; 47322 MW; 2563589957FAAD64 CRC64;				
Query Match	27.4%; Score 611.5; DB 3; Length 426;				
Best Local Similarity	35.2%; Pred. No. 5.8e-36;				
Matches	153; Conservative 71; Mismatches 144; Indels 67; Gaps 13;				
QY	1 MSKAPSGKAFKLAGYIVLPEITDAYVTALSSPMHAYITPKGSLKSRKISSPPFAN 60				
DB	1 MKVTCSPGKVLIAAGYIVLDPQYSGLVITGLTAGYASTTLTD--KCGTRVKSPPQIN 58				
QY	61 GEMVHYSNTEKR-----EVQSRINPELATITFIYALYI-----QTEAFDLEIT 107				
DB	59 AEWLYNIDWTVPPIRVHQLIYENCLEKNPNEVQLAFYVINYFPFGTGRQLCQMDLQVT 118				
QY	108 IYSDPGYHSQEDTEKTSNGEKTFLYHSRAITEVEKTKGLSSAGLVSVAATSL--SHFI 166				

```

Db      119  LQVDAAYHNGPOLKEDQGTSPKXENFL--NCTLQGVHKTGLGSSAAMI-----TSLIGSLFL 172
QY      167  P-----NVISTKNDILHNVAOIAHCYAKKIGSGFDVATAIGLIVRRPOP 213
Db      173  SLRLRLDDTGDGSLKIDIDSTKIVVHNLAIHCSAGKVGSGFDVGAATGSCIVRRFDP 232
QY      214  ALINVPQVLBSDEP--KEPTKIKLIESNMEKHKRCLPYIK---LLMDVVGSGE 267
Db      233  KLIEQLLPYDEQIINKININFTSRKIVSKKMSD---VVPFLPATYICLLMGDVAAGSS 287
QY      268  TPKLSRYLQWKKKEKPEESSVYVDOLNSANLOFMKELREMRREKXDSDPEYIIEKLDHAYE 327
Db      288  TPGMKRKYQOMKXENPEESKNCPEDDLYSKVLS--IKKCPILSSBLSDE----- 333
QY      328  PLTVAIKNIRKGLQALTKQSEVPIEBEDVOTQLDLDRQCEIIPGCVGVVPGAAGYDALAVL-- 386
Db      334  -LQSGFRSIRRLIGRITVEAKVDIEPLTKQTNILNIEQLPGVIGVGVPGAAGGDAFCCLA 392
QY      387  -----VLENOGVNFK 396
Db      393  INHTEIENVIKTKW 407

```

RESULT 2	ID	Q9C6T1	PRELIMINARY;	PRF;	505 AA.
AC	Q9C6T1				
DC	Q9C6T1				
DT	01-JUN-2001	(TREMBLrel. 17, Created)			
D7	01-JUN-2001	(TREMBLrel. 17, Last sequence update)			
D7	01-OCT-2002	(TREMBLrel. 22, Last annotation update)			
DE	Hypothetical 54.4 kDa protein.				
GN	FSM6.9.				
OC	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsis.				
NCBI	Taxid=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=cv Columbia;				
EX	MEDLINE=21016719; PubMed=11130712;				
RA	Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,				
RA	White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,				
RA	Buhler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,				
RA	Chung M.K., Conn L., Conway A.R., Conway A.R., Creasy T.H., Dewar K.,				
RA	Dunn P., Etgu P., Feldhym T.V., Feng J.-D., Fong B., Fujii C.Y.,				
RA	Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huzar L.				
RA	Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,				
RA	Kim C.J., Koo H.L., Kremetschka I., Kurtz D.B., Kwan A., Lam B.,				
RA	Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,				
RA	Lin X., Liu S.X., Liu Z.A., Luos J.S., Maiti R., Marziani A.,				
RA	Miltchev J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,				
RA	Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,				
RA	Sakano H., Salazar S.L., Schwartz J.R., Shim P., Southwick A.M.,				
RA	Sun H., Tallon L.J., Tambunga G., Toriumi M.T., Town C.D.,				
RA	Utechtack T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,				
RA	Wu D., Yu G., Frazer C.M., Venter J.C., Davis R.W.;				
RT	"Sequence and analysis of chromosome 1 of the plant Arabidopsis				
RT	thaliana.";				
RL	Nature 408:816-820(2000).				
DR	EMBL; AC079041; AAG50716.1; -.				
DR	InterPro; IPR005916; Pnev_kin_eng8.				
DR	TIGRFAMS; TIGR01219; Pnev_kin_ER08.1.				
kw	Hypothetical protein.				
Q0	SEQUENCE 505 AA; 54409 MW; BICBA6CA338BD63 CRC64;				

Query Match	21.3%	Score 475;	DB 10;	Length 505;
Best Local Similarity	28.3%;	Pred. No. 5.3e-26;		
Matches 143; Conservative	78;	Mismatches 184;	Indels 100;	Gaps 15
QY	1	MSKAFSAPGKFLAGGYLVLEPTIDYAYVTTALSSRNHAYTPKGTSLKE-----SRIKI	53	

```

Db      MAVASAPGVMTGTGYLVLEKPNMGLVLTSTNARFYAIVKDINEBVRKESWAMKMTDVKL 60
QY      54  SSPOFANGEMEHIISN--TEKPREVQSRINPFLBATTPIYLAITYOPLEAD----- 105
Db      61  TSPQLSR-BSMYGLSINHLTLTQSVASADSRRNPFEMHAIQYAAIAAHLATEKDKESHLKL 119
QY      104 ---LEIITYSDPGYHQED-----TEKTSNGEKTFLYHSR 137
Db      120 LQGLDITILGSDNPFYSYRNQIESAGLPLTPESLGTLPAPASITTNAAESNG-----AN 177
QY      138 AITEVEKTLGSSAGLVSVATSLSHFTPNYIS-----TNKDIILNVAQIAH 185
Db      173 SKREVAKTGIGSSPAATTAIVVAALL-HYL-GVWLDSPCKEKGKGCSPDLVTHIMIAQTSH 230
QY      186 CYAKOKIGSGFDVAATLYGLIYRRFPQPLINDVYQVLESPP--EKPFTELKULIESNWE 243
Db      231 CLAQKVGSGFDVSCAVYSGRYRFAFSEBVLSPAOAVATGLPLNBEVGTILK---GKMD 268
QY      244 EKHERCTLPYGIKLMGDV-KGSGSETPKLVBSHVLQMKKEKRESSVYVDQINSANLQPMK 302
Db      287 NKRTESPFLPMLFLGEBPGSGSSSTPSVGAVKMKQMSDEKREMNQUNSLDANLELT 346
QY      303 ELKEMEKXDSDETYIKELD-----HSVEPLTYAI-----KNIRKG 339
Db      347 KUNDLSLAKOHMDVLRVTKCSVLTSEKWTLAHTSEITNEBAILIKELIARBEAMLRIRIL 406
QY      340 LQALTKSEVPIEPDQOTLLDRCOEIPGCVGVPPGAGYDAIAYVLNMQVNPFQKT 398
Db      407 MRQMGASAVPIEPESQOTLLDSTMSAGVILLAGVPAGGFDALFALTLDGSGTKLQWAM 466
QY      400 LBNPDYFNHYVWVDEBQESGVLEE 424
Db      467 SS-----HNVLALVNEDEHGVGLE 486

```

RESULT 3	0944G1	PRELIMINARY;	PRT;	503 AA.
ID	0944G1			
AC	0944G1;			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)			
DE	Phosphomevalonate kinase.			
OS	Hevea brasiliensis (Para rubber tree).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid I; Malpighiales; Euphorbiaceae; Hevea.			
OX	NCBI_TaxID=3981;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. KRM600;			
RA	Hallahan D.L., Keiper-Hrynko N.M.;			
RT	"Genes involved in the biosynthesis of isopentenyl diphosphate in the rubber tree <i>Hevea brasiliensis</i> ."			
RL	Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AF429385; AAL18926.1; -			
DR	InterPro; IPR005916; Pmev_kin_ei88.			
DR	TIGRFAMs; TIGR01219; Pmev_kin_ER88. 1.			
KW	kinase.			
SQ	SEQUENCE 503 AA; 54171 MW; BE6F91B80B45FF94 CRC64;			
Query Match	20.0%;	Score 445.5;	DB 10;	Length 503;
Best Local Similarity	27.4%;	Pred. No. 7.1e-24;		
Matches 138;	Conservative 88;	Mismatches 179;	Indels 99;	Gaps 16

[illegible]

```

Qy 104 ----LEIIYSDPGYHSGED-----TETKTSNGEKTFLYHS 136
Db 120 LLOGIDITLGTNDYSYRNEIEACGLPLTPRESIALPFSISITFNVEBANQO----- 172
Qy 137 RAITEVEKGTGSSAGLVSVAATSLSHF-IPNVISTNK-----DILHNVAQIAHCYA 188
Db 173 NCKPEVAKTGTGSSAMTAATVAALHHLGLVDLSSSCCKEKKSPDLVHIIAQAHGIA 232
Qy 189 OKKIGSGFDVAATAYGLIVRRFOPALINDVYQVLESDEPEKPTBLKLIKSNEEKKER 248
Db 223 QGKVGSGFDVSSAVYGSRRYRFPSEVLSSADACKGIP--LQEVISNLIKMKHEKTM 290
Qy 249 CTLPGYIKTLMGDV-KGSGSETPKLVSRVQMKKEKPESSVYDQNSAN-----IQFM 301
Db 291 FSLPPLMELLIGPEGTGSSTPSMGALKMKQSDTKQSQETWRKLSBANSALETOFNIL 350
Qy 302 KEIEMREKRYDS-----DEETIKELDHSVEPLTVAI-----KNIRKGLQA 342
Db 351 SKLAE--EHMDYKVCVIDSCSTKSEKWEIQATEPSREAVVALLGSRNMLQIRNVRQ 408
Qy 343 LTKSEVEPIEPDVOTQLDRCOEIPGCVGVPGAGYDATALAVLENOVGNFKQ--KTL 400
Db 409 MGEAAGVIEPIEBSQRLDDTTMNMGVLLAGVPGAGFDAYPAVTLDGSGTNVAKAMSL 468
Qy 401 ENPDYFHNWYWDLEEQTEGVLEB 424
Db 469 -----NVLLALVREDPENGVLLE 485

```

RESULT 4

```

Q8ETN1 ID Q8ETN1 PRELIMINARY; PRT; 369 AA.
AC Q8ETN1;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Phosphomevalonate kinase.
GN OBO227.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 11309;
RX MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments."
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AP004593; BAC12183.1;
KW Kinase; Complete proteome.
SQ SEQUENCE 369 AA; 40440 MW; ABBF89515CSF3E2 CRC64;

```

Query Match 11.1%; Score 247; DB 16; Length 369;
 Best Local Similarity 23.4%; Pred. No. 9.8e-10;
 Matches 107; Conservative 67; Mismatches 162; Indels 122; Gaps 21;

```

Qy 1 MSK---AFSAGKFLAGGYVLEPIYAYTALSSRMHVAITPKGSLKESRIKISPO 57
Db 1 MSKQQLTVKVGKLMIGEPFAVLEBYOKLAWAVRVYATIT-----DAENYTLIED 54
Qy 58 FANGE--WEYHIS--NTEKPREVOSRINPFLAETFIYLAIOPTAEADLEIIISDPGY 114
Db 55 FGLHDIAMSYOTGVNIHSDERS---FVKDAIYITCYLLEKS-----IGLPSF 102
Qy 115 HSQEDTETKTSNGEKTFLYHSRAITEVEKGTGSSAGLVSVAATSLSHRIPNVISTNK 174
Db 103 HLSTRSELDASG-----VKYGLSSAIVTSVSAIRFAPARY--ATK 145
Qy 175 DILHNVAQIAHCYAOKKIGSGFDVATAIY-GLIYRRFOPALINDVFOVLESDEKPEPTE 233
Db 146 EIRFLALSHVCTQGN-GSGADIAASTYGGFRQYSSFOADWLNTY-----KNAST 196

```

```

Qy 234 LKLIESNWE-EKERCTLPGYIKTLMGDVKGSGSETPKLVSRVQMK-----KE 281
Db 197 ISEVYDMEKRYQGENIRIPDLFNVAVCGWGTCTAATSKTLVQGIHLKXSDDAYEHFIQS 256
Qy 282 KPRESSVYDQNSANIQFMKELREMBRKYSDDETYIKELDHSVEPLTVAIKNIRKGLQ 341
Db 257 SKEAVGLFLDDMNSENLSMLMD-----GVRKNRQ 285
Qy 342 ALT---OKSEVPIEPDVOTQLDRCOEIPGCVGVPGAGYD-AIANTVLENOVGNFKQ 397
Db 286 ALAHVQQAQKPTITSMULTKLCDAEDUG--AKPSGAGGDCGIAPMYSKDAQENL-- 341
Qy 398 KTLNPDYFHNWYWD-----LEQTEG---VLEEK 425
Db 342 -----FH--AAEFGIKPLTIQPSLESGQVIEEB 369

```

RESULT 5

```

Q973B5 ID Q973B5 PRELIMINARY; PRT; 314 AA.
AC Q973B5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein ST0978.
GN ST0978.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OX Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Anka A., Koenig H., Hosoyma A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamitsu M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermocacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7."
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000984; BAB65998.1;
DR InterPro; IPR006204; GMP_kinase.
DR Pfam; PF00286; GMP_kinases; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 314 AA; 35307 MW; 89740213AB16335F CRC64;

```

Query Match 10.2%; Score 228; DB 17; Length 314;
 Best Local Similarity 22.6%; Pred. No. 1.8e-08;
 Matches 98; Conservative 71; Mismatches 121; Indels 144; Gaps 18;

```

Qy 6 SAPGKFLAGGYVLEPIYD--AYVTALSSRM-----HAVITPKGSLKE 48
Db 3 SAPKILMIGSYIS---VPGISHVIAINKVRCDIYSSNNFIRETYGFPKXGNELIE 59
Qy 49 SRKISSPQFANGGEWEYHISNTEKPREVOSRINPFLAETFIYLAIOPTAEADLEII 108
Db 60 SVITVFKEKGS-----LPPF-----HYKL 79
Qy 109 YSDPGYHSGEDTETKTSNGEKTFLYHSRAITEVEKGTGSSAGLVSVAATSLSHRIPN 168
Db 80 FND-----KQFIHGK-----KTLGSSSSASTALPACTIYYIYLPKN 115
Qy 169 VISTNKDILHNVAQIAHCYAOKKIGSGFDVATAIYGLIYRRFOPALINDVFOVLESDEPE 228
Db 116 L---NKOETIKLAQKANYIKQKIGSGFDIASAVGSIYVRRF-----DIE 159
Qy 229 KPFTLKLKLIESNWEKHERCTLPGYIKTLMGDVKGSGSETPKLVSRVQMKKEKPESSV 288
Db 160 KVDVIEPLKIGNYE-----MLGFIAGESFTVNSVAKPIE--KSNNEBFKK 204

```

```

OY 289 VYDOLNSANTLOFMKELREKREKYSDDPEFYIKELDHSVEPLTVAIKNIRKGLQALTKQ-S 347
DB 205 VAKYIDEEINMAIKLIK-----LGIKEIAIHVKLA-----RRFLANGLAKKIV 247
OY 348 EVPIEDVQOTLLDRCEIPGCGVVPAGGYDAIAVLVLNQNQGNFKQKTLNPDYH 407
DB 248 GVEIENEKIRRLIEMEN--DALIALSPGAGS---SVFLGDKLSKYEKEK-----E 297
OY 408 NYVWDLBEQTEGV 421
DB 298 NVIVIELKED-EGT 310

RESULT 6
ID 097UL6 PRELIMINARY; PRT; 323 AA.
AC 097UL6;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE (Phospho) mevalonate kinase, putative.
GN SSO2988.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtiss B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K., Medina N., Peng X.,
RA Helkamp de Jong I., Jeffries A.C., Kozera C.J., Theriault C., Tolstrup N.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RA RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RA Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840 (2001).
DR EMBL; AB006889; AAK43093.1; -.
DR InterPro; IPR001174; Galkinase.
DR Pfam; PF00288; GHMP_kinases; 1.
DR PRINTS; PR00960; LMBPPOTEIN.
DR KINASE; Complete proteome.
SQ SEQUENCE 323 AA; 36207 MW; D62B48552P5CDBF8 CRC64;

Query Match 10.1%; Score 225; DB 17; Length 323;
Best Local Similarity 23.7%; Pred. No. 3, 1e-08;
Matches 100; Conservative 67; Mismatches 141; Indels 114; Gaps 16;

OY 6 SAPGKAFLAGYLLEPIYDAYVTALSSRMHAIVTPKGTSLKESRIKISSPOFANGEMEX 65
DB 5 SAPGKILWNGSYV-----FGISHVIAVNRKVCGLREIKEDSLIFHTSYGH 54
OY 66 HISSMTEKREYQSRINPFLTEITIVLAAYIOPTAFDLEIIITSDPGHSEDEETKTS 125
DB 55 FKNSSNEL--INSVLDTFRRLSQLPGY-----EIDLVD----- 88
OY 126 SNGEKTFLYHSRAITEVEKTLGSSAGLVAVATSLSHFIPNVISTNKDI--LHNVAQI 183
DB 89 ----KEF-----IDGKKTGLSSS-----AATSLTACLYVALHKGKLDLEIHKLAQI 133
OY 184 AHCVAKKIGSGFDVATAIYGLIYRRFOPALINDVFOVLESDEPEKPTLKLIESNWE 243
DB 134 ANYKQKIGSGFDIASAVFGSIVYKRF-----TDLKMPFYFEKLMIGNVD 180
OY 244 EKHGECTLPYGIKILMGDVKGSSEPKVSRVLQMKKEKRESSVYDOLNANLOFMKE 303
DB 181 -----MMLGFTKGSSEVTVGLVRKFE-----KSNLDPEE 210
OY 304 IREMEKYSDDPEFYIKELDHSVEPLTVAIKNIRKGLQALTKQSE---VPIEDVQOTL 359

```

```

DB 211 IMRL-----IDENVMAIKLIKLNKLDVAEVIKLGKRYLVIAERIYGVVLVSGNEBEL 265
OY 360 LDRCEIPGCGVVPAGGYDAIAVLVLNQNQGNFKQKTLNPDYHNVWDLBEQTE 419
DB 266 IKIAE--BGALVALSPGAGGDSI--FALGNDLRNREAWSK-----RGIFIDVKED-E 316
OY 420 GV 421
DB 317 GL 318

RESULT 7
ID 09FD67 PRELIMINARY; PRT; 368 AA.
AC 09FD67;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Phosphomevalonate kinase.
GN MVAK2.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20353468; PubMed=10894743;
RA Wilding E.T., Brown J.R., Bryant A.P., Chalke A.F., Holmes D.J.,
RA Ingraham K.A., Jordanescu S., So C.Y., Rosenberg M., Gwynn M.N.;
RA "Identification, evolution and essentiality of the mevalonate pathway
RA for isopentenyl diphosphate biosynthesis in gram-positive cocci.";
RA J. Bacteriol. 182:4319-4327 (2000).
DR EMBL; AF290093; AAG02442.1; -.
DR InterPro; IPR006204; GHMP_kinase.
DR InterPro; IPR006206; Mew_galkinase.
DR InterPro; IPR005917; Pnev_kin_Gpos.
DR Pfam; PF00288; GHMP_kinases; 1.
DR PRINTS; PR00959; MEVGALKINASE.
DR TIGRFAMs; TIGR01220; Pnev_kin_Gr_pos; 1.
DR KINASE.
SQ SEQUENCE 368 AA; 40541 MW; 390993D36577146D CRC64;

Query Match 9.4%; Score 209.5; DB 2; Length 368;
Best Local Similarity 27.0%; Pred. No. 5e-07;
Matches 113; Conservative 52; Mismatches 137; Indels 117; Gaps 23;

OY 6 SAPGKAFLAGYLLEPIYDAYVTALSSRMHAIVTPKGTSLKESRIKISSPOFA----- 59
DB 5 TTPGKLFAGHYAVVEPCHPAIIAVD--QEVTVVEETDEGS---IQSAQYSLPIRW 59
OY 60 --NGEMETHSSNTEKREYQSRINPFLTEITIVLAAYIOPTAFDLEIIITSDPGHYS 116
DB 60 TRRNGELVLDI-----RENPF-----HYVLAAILHTEKY-----A 89
OY 117 QEDTEKTSNGEKTFLYHSRAITEV-----KTGLSSAGLVAVATSLSHFIPNVIS 171
DB 90 DE-----QKELSF-YHAKVTSELDSSNKRKTGLSSS-GAVYGVYKALNIFID--LG 138
OY 172 TNKDIHLNVAQIAHCYAKKIGSGFDVATAIY-GLIYRRFOPALINDVFOVLESDEPEK 230
DB 139 LENEFIFPLSALHIAVQGN--GSCGDIAASCYGMIAFSTPDHVMVN-----QKV 187
OY 231 PTE-LKTLIESNWE-----KHRECTLPIYGIKILMGDVKGSSEPKVSRVLQMKKEK 283
DB 188 TTEFTLTDLAWDELMIFPLK-----VPKQRLILIGWTGSPASTSDLVNDVHOSKEEK- 241
OY 284 EESSVYDOLNSANTLOFMKELREKREKYSDDPEFYIKELDHSVEPLTVAIKNIRKGLQAL 343
DB 242 -----QAAVEQFLMKRLC-----VETMIN--GENTKISIVYQKQITRNQL 282
OY 344 TOKSE--VPIEDVQOTLLDRCEIPGCGVVPAGGYDAIAVLVLNQNQGNFKQKT 399
DB 283 AELSSLTGVVLETALKNLCDLASVYGAAS--SGAGGGDCGIVI-----FRQGS 331

```

RESULT 8
Q9FD77 PRELIMINARY; PRT; 358 AA.
AC Q9FD77; 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Phosphomevalonate kinase.
GN MVAK2
OS Staphylococcus haemolyticus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1283;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20353468; PubMed=10894743;
RA Wilding E.I., Brown J.R., Bryant A.P., Chalker A.F., Holmes D.J.,
Ingraham K.A., Iordanescu S., So C.Y., Rosenberg M., Gwynn M.N.;
RT "Identification, evolution and essentiality of the mevalonate pathway
for isopentenyl diphosphate biosynthesis in gram-positive cocci.";
RL J. Bacteriol. 182:4319-4327(2000).
DR EMBL; AF290089; AAC02432.1; -
DR InterPro; IPR006204; GMP_Kinase.
DR InterPro; IPR006206; Mgv_Galkinase.
DR InterPro; IPR005917; Pmev_Kin_Gpos.
DR Pfam; PF00288; GMP_kinases; I.
DR PRINTS; PRO0959; MEVGALKINASE.
DR TIGRFAMs; TIGR01220; Pmev_kin_Gr_pos; 1.
KM Kinase.
SQ SEQUENCE 358 AA; 39908 MW; C31755A4D57A33A CRC64;

Query Match 8.7%; Score 193; DB 2; Length 358;
Best Local Similarity 24.0%; Pred. No. 7.4e-06;
Matches 101; Conservative 63; Mismatches 129; Indels 128; Gaps 21;
QY 7 AFGKALAGGYLVLEPIY-----DAYVTLSSRMHVI-----PKGTSLKES 49
DB 6 AFGKLYAGEYAVTEPGKSVLIANDRFVTASIEASNAVSTIHSKTLHYEPVFNRED 65
QY 50 RIKISSPOFANGEMEYHISSTKREXVOSRINPLEATPIVLAYIOPTFAFDEIITY 109
DB 66 KIDISDANAASQ-----LKVVTALIEVFEQYVASCNVKL--- 99
QY 110 SDPGYHSOEDTETKTSNGEKTFLYHSRAITEVEKTLGSSAG--LVSVATSLSHPIFN 168
DB 100 --KHFHLEIDSUDDASG-----NKGIGSSAAVLAVSV--KALNEFYDM 140
QY 169 VISTNKDILHNVAQIAHCYAQKIGSGFDVATAY--GLIVRRFPQALIND--VFQVLES 226
DB 141 QLSNL--YIKLAVISNRQLSSCG--DIAVSYSGLAVSTFD---HDMVQOMEE- 192
QY 227 PEKPTTELKLIIESWEKX-ERCTLPYGIKLMGDVKGSETPKLVSRVLQWKKEPKEE 285
DB 193 -----TSVNEVLEKMWPGHIEPILOAPENMEVLIGWTGSPASSPVLVSEVRKL- 241
QY 286 SSVVYDQNSANLQFMKELREMEKXSDPEYIKELDHS---VEPLTVAIK--NIRKGL 340
DB 242 -----SDPSFYGDPLDOSHTCVENLLYARFTDTI-KGV 273
QY 341 QALQKSEVPIEP-----DVQTLDRCOEIPGCVGV--VPAGAGYDAIAVLVLENO 391
DB 274 QKMLRQNMITIQOMDNEATVQIETENLMLQCDIGERYAAAKTSAGAGGDC-GIAIINDR 332
QY 392 V 392
DB 333 I 333

RESULT 9
Q9FD72 PRELIMINARY; PRT; 358 AA.
ID Q9FD72;
AC Q9FD72;

DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Phosphomevalonate kinase.
GN MVAK2 OR SE0363.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20353468; PubMed=10894743;
RA Wilding E.I., Brown J.R., Bryant A.P., Chalker A.F., Holmes D.J.,
Ingraham K.A., Iordanescu S., So C.Y., Rosenberg M., Gwynn M.N.;
RT "Identification, evolution and essentiality of the mevalonate pathway
for isopentenyl diphosphate biosynthesis in gram-positive cocci.";
RL J. Bacteriol. 182:4319-4327(2000).
DR EMBL; AF290091; AAC02437.1; -
DR EMBL; AE016745; AAC03960.1; -
DR InterPro; IPR006204; GMP_Kinase.
DR InterPro; IPR006206; Mgv_Galkinase.
DR InterPro; IPR005917; Pmev_Kin_Gpos.
DR Pfam; PF00288; GMP_kinases; I.
DR PRINTS; PRO0959; MEVGALKINASE.
DR TIGRFAMs; TIGR01220; Pmev_kin_Gr_pos; 1.
KM Kinase; Complete proteome.
SQ SEQUENCE 358 AA; 39886 MW; 608EBD4D5774EC2 CRC64;

Query Match 8.1%; Score 181.5; DB 16; Length 358;
Best Local Similarity 24.1%; Pred. No. 5e-05;
Matches 99; Conservative 52; Mismatches 131; Indels 129; Gaps 18;
QY 7 AFGKALAGGYLVLEPIYDAYVTLSSRMHVI-----TPKGTSLKES 49
DB 6 AFGKLYAGEYAVTEPGKSVLIANDRFVTASIEASNAVSTIHSKTLHYEPVFNRED 65
QY 50 RIKISSPOFANGEMEYHISSTKREXVOSRINPLEATPIVLAYIOPTFAFDEIITY 109
DB 66 RIEISDVQAQ-----LKVVTALIEVFEQYVASCNVKL--- 99
QY 110 SDPGYHSOEDTETKTSNGEKTFLYHSRAITEVEKTLGSSAG--LVSVATSLSHPIFN 168
DB 100 --KHFHLEIDSUDDASG-----NKGIGSSAAVLAVSV--KALNEFYDM 139
QY 169 VISTNKDILHNVAQIAHCYAQKIGSGFDVATAY--GLIVRRFPQALIND--VFQVLES 226
DB 140 LELSNLYIKLAVISNRQLSSCG--DIAVSYSGLAVSTFD---HDMVQOMEE- 192
QY 227 PEKPTTELKLIIESWEKX-ERCTLPYGIKLMGDVKGSETPKLVSRVLQWKKEPKEE 285
DB 193 -----TSVNDVLEKMWPGHIEPILOAPENMEVLIGWTGSPASSPVLVSEVRKL- 241
QY 286 SSVVYDQNSANLQFMKELREMEKXSDPEYIKELDHS---VEPLTVAIK--NI----- 336
DB 242 -----SDPSFYGDPLDOSHTCVENLLYARFTDTI-KGV 274
QY 337 -----RKGLQALQKSEVPIEPDVQTLDRCOEIPGCVGV--VPAGAGYDAIAVLVLENO 381
DB 275 KMIRINRRIITQOMDNEASVLEITDKLKLCDVGEKGG--ASKTSAGAGGD 323

RESULT 10
Q8NXXO PRELIMINARY; PRT; 358 AA.
ID Q8NXXO;
AC Q8NXXO;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

RESULT	11
Q9FD62	
ID	Q9FD62
AC	Q9FD62;
DT	01-MAR-2001 (TREMBLrel. 16, Created)
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE	Phosphomevalonate kinase.
GN	MVAK2.
OS	Enterococcus faecium (Streptococcus faecium).
OC	Bacteria; Firmicutes; Lactobacilliales; Enterococcaceae; Enterococcus
NCBI	taxid=1352;

```

RESULT 12
09FD83
ID 09FD83 PRELIMINARY; PART; 358 AA.
AC 09FD83
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Phosphomevalonate kinase.
GN MVAK2.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcu.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20353469; PubMed=10894743;
RA Wilding E.I., Brown U.R., Bryant A.P., Chalker A.F., Holmes D.J.,
RA Ingraham K.A., Iordanescu S., So C.Y., Rosenberg M., Gaym M.N.;
RT "Identification, evolution and essentiality of the mevalonate pathway
RT for isopentenyl diphosphate biosynthesis in gram-positive cocci.";
RL J. Bacteriol. 182:4319-4327(2000).

```

Query Match	7.8%;	Score 173;	DB 2;	Length 358;
Best Local Similarity	22.8%;	Pred No. 0.00021;		
Matches	97;	Conservative	57;	Mismatches 143; Indels 128; Gaps 20

RESULT 13

RT aureus.";

Query Match	7.8%:	Score 173;	DB 16;	Length 358;
Best Local Similarity	22.8%:	Pred. No. 0.00021;		
Matches 97;	Conservative 57;	Mismatches 143;	Indels 128;	Gaps 20

RESULT 14

RA Mañueno E

RA Nordstiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.,
 RT "Comparative genomics of *Listeria species*,"
 Science 294:849-852(2001).
 DR EMBL; AL591973; CAC98227.1; -.
 DR ListerList; LMO00012; -.
 DR InterPro; IPR006204; GHMP_kinase.
 DR InterPro; IPR006206; Mey_galkinase.
 DR InterPro; IPR005917; Pmev_kin Gpos.
 DR Pfam; PF00288; GHMP_kinases; I.
 DR PRFAMS; PR00959; MEVGALKINASE.
 DR TIGRPMs; TIGR01220; Pmev_kin Gr_pos; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 359 AA; 39981 MW; 15156CA0BD97A1C CRC64;

Query Match 7.6%; Score 170.5; DB 16; Length 359;
 Best Local Similarity 24.7%; Pred. No. 0.00031;
 Matches 96; Conservative 55; Mismatches 144; Indels 93; Gaps 19;

QY 8 PGKAFLAGYLVLEPIYDAVYVTAALSRMHAIVTPKGTSLKESRIKISSPOFANGEMEYHI 67
 DB 13 PGKLYAGAYAVESGHTAILTR---VNRIT--LTLEDS-----RNEIMTPHY 57
 QY 68 SNTKEPREVOSRINP-----FLEATIFIVAYIOTPEAPDL---EIIYSDPGYHSOE 118
 DB 58 ENPVSWP--VGELKPDGHEMTFTAENIATTFELK--SEGIELTPVKMI----- 104
 QY 119 DTEFTSSNGEKTFLYHRAITVEKGTGSSAGLVSVATSLSHFIPNVIISTKNDILH 178
 DB 105 ETELIDOGSA-----KYGLSSAAATVAVINAMLTFFPEISMLK--F 146
 QY 179 NVAQIAHCYAOKKIGSFDAVTAIY-GLIVRRFOPALINDVFOVLESDEKPEFTELK 237
 DB 147 KLAALSHLVQGN-GSCGDIASCMYGMIAITTFDQEWK--HRLAYKSLFMKEPPPM 203
 QY 238 IESNNEEKHERCTLPYGIKILMGDVKGSETP---KLVSRYLQWKKKEPBESSVYDOL 293
 DB 204 LQI---ETLEBPVPFVS-----GWTGTPVSTGKLVSQIHAFK---QEDSKNYOHF 248
 QY 294 NSANIQFMKELREKRDSDPEYIKELDHSVEPLTVAIKIRKLOALTOKSEVPPIEP 353
 DB 249 LTRNNEIMKOIIQAFTTDE-----ELLYSAIKENRRILQELGTAGVNIET 295
 QY 354 DVQTOILDRCOEIPGCVGAVPGAGYD 381
 DB 296 SLIKELADSAENMG--AGKSSGSGGD 321

RESULT 15

092FU1 PRELIMINARY; PRT; 360 AA.
 AC 092FU1; 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hypothetical protein lin0012.
 GN LIN0012.
 OS *Listeria innocua*.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 OX NCBI_TaxID=1642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIP 11262 / Serovar 6a;
 RX PubMed=11679669;
 RA Glauber P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Chablit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,
 RA Entian K.-D., Fehli H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hahn T., Haut J., Jackson D.,
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkak G.,
 RA Madueno E., Maltournam A., Mata Vicente J., Ng E., Nedjari H.,

RA Nordstiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.,
 RT "Comparative genomics of *Listeria species*,"
 Science 294:849-852(2001).
 DR EMBL; AL596163; CAC95245.1; -.
 DR ListerList; LIN00012; -.
 DR InterPro; IPR006204; GHMP_kinase.
 DR InterPro; IPR006206; Mey_galkinase.
 DR InterPro; IPR005917; Pmev_kin Gpos.
 DR Pfam; PF00288; GHMP_kinases; I.
 DR PRFAMS; PR00959; MEVGALKINASE.
 DR TIGRPMs; TIGR01220; Pmev_kin Gr_pos; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 360 AA; 40110 MW; 0FFDD7B062B42584 CRC64;

Query Match 7.4%; Score 165.5; DB 16; Length 360;
 Best Local Similarity 24.0%; Pred. No. 0.00073;
 Matches 93; Conservative 58; Mismatches 144; Indels 93; Gaps 19;

QY 8 PGKAFLAGYLVLEPIYDAVYVTAALSRMHAIVTPKGTSLKESRIKISSPOFANGEMEYHI 67
 DB 14 PGKLYAGAYAVESGHTAILTR---VNRIT--LTLEDS-----RNEIMTPHY 58
 QY 68 SNTKEPREVOSRINP-----FLEATIFIVAYIOTPEAPDL---EIIYSDPGYHSOE 118
 DB 59 ENPVSWP--IGELKPDGHEMTFTAENIATTFELK--SEGIELTPVKMI----- 105
 QY 119 DTEFTSSNGEKTFLYHRAITVEKGTGSSAGLVSVATSLSHFIPNVIISTKNDILH 178
 DB 106 ETELIDOGSA-----KYGLSSAAATVAVINAMLTFFPEISMLK--F 147
 QY 179 NVAQIAHCYAOKKIGSFDAVTAIY-GLIVRRFOPALINDVFOVLESDEKPEFTELK 237
 DB 148 KLAALSHLVQGN-GSCGDIASCMYGMIAITTFDQEWK--HRLAYKSLFMKEPPPM 204
 QY 238 IESNNEEKHERCTLPYGIKILMGDVKGSETP---KLVSRYLQWKKKEPBESSVYDOL 293
 DB 205 LQI---ETLEBPVPFVS-----GWTGTPVSTGKLVSQIHAFK---QEDSKNYOHF 249
 QY 294 NSANIQFMKELREKRDSDPEYIKELDHSVEPLTVAIKIRKLOALTOKSEVPPIEP 353
 DB 250 LTRNNEIMKOIIQAFTTDE-----ELLYSAIKENRRILQELGTAGVNIET 296
 QY 354 DVQTOILDRCOEIPGCVGAVPGAGYD 381
 DB 297 SLIKELADSAENMG--AGKSSGSGGD 322

RESULT 16

09KM3 PRELIMINARY; PRT; 374 AA.
 AC 09KM3; 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Phosphomevalonate kinase.
 GN Streptomyces sp. (strain CL190).
 OS Streptomyces sp. (strain CL190).
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 OC Streptomyces; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=93372;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CL190;
 RX MEDLINE=20353446; PubMed=10894721;
 RA Takagi M., Kuzuyama T., Takahashi S., Seto H.,
 RT "A gene cluster for the mevalonate pathway from *Streptomyces* sp.
 strain CL190,"
 RL J. Bacteriol. 182:4153-4157(2000).
 DR EMBL; AB037666; BAB07792.1; -.
 DR InterPro; IPR001969; Aspproteinase site.
 DR InterPro; IPR006204; GHMP_kinase.
 DR InterPro; IPR006206; Mey_galkinase.

DR InterPro; IPR005917; Pmev_kin_Gpos.
 DR Pfam; PF00288; GHMP_kinases; 1.
 DR PRINTS; PR00959; MEVGALKINASE.
 DR TIGRFAMs; TIGR01220; Pmev_kin_Gr_pos; 1.
 DR PROSITE; PS00141; ASP_PROTEASE; 1.
 KW kinase

SO SEQUENCE 374 AA; 39281 MW; A2908B333C694E6D CRC64;

Query Match 6.6%; Score 147; DB 2; Length 374;
 Best Local Similarity 22.4%; Pred. No. 0.017;
 Matches 100; Conservative 61; Mismatches 161; Indels 124; Gaps 23;

QY 7 AAGKAFLAGYVLEPIYDAYVATLSSRMHVAITPKGSLKSRKISS---PQFANGEM 63
 DB 12 AAGKLFVAGEYAVVDPGNPAIIIVAVDRHISVTYSDADADTGAADVSSDLGQAVG--W 69
 QY 64 EKH-----ISSNTEKREVSRIINPLEATIFIVLAVIOPTFADEIITISDPGHISQD 119
 DB 70 RWHDELIVRDPDDGQARSALAHVVSAT-----ET 100
 QY 120 TETKTSNGEK-----TFYHSRAITEVEKTKGSSAGLVSVVATSLSHFIPNVISTNKD 175
 DB 101 VGRLLGERGQKYPALTLVSSSRHEDGRKFGLGSS--GAVTVATVAANAFAFCGLEIST--D 157
 QY 176 ILHNAVAQIAHCAQAQKIGSGFVATAIV--GLIVRRFPQALINDVFOVLESDEPKFTEL 234
 DB 158 ERFRLAMLATIALDPRK--GSGGDLAASTWGMIAYQAPDRAFYDLARVGV----- 208
 QY 235 KKLISNNR--EKHERCTLPYGIKLMGDVKGSETPKIVSRVLO--WKEKEESSY--V 289
 DB 209 -RTLPAWPGHSHVRRLPAPKGLTLEVGWTEGPASTASLSDIHRRTWRGSASHQRFVET 267
 QY 290 YDOLNSA-----NLQFMKELREWEKDYSDPEYIKELDSVEPLTVAIKIRKGLQ 341
 DB 268 TDCVSAVAVALSSGDDTSLHIRARQD-----LARDDEV-----GLG 307
 QY 342 ALYQSEVPIEPDVOTQLDRQCEIPGCVGVV--FGAGGYD-AIAVYLE--NOVGNFK 396
 DB 308 IFTPK-----LTALCDABE---AVGGAAGSGAGGCGGALALDAEASRDIITHR 354
 QY 397 QKTLENPDYFHNVYVVDLEEQTEGYL 422
 DB 355 QR-----W-----ETAGVYL 363

RESULT 17

Q93K31 PRELIMINARY; PRT; 294 AA.

AC Q93K31;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Phosphomevalonate kinase.
 GN PMK.
 OS Lactobacillus helveticus.
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
 OC Lactobacillus.
 OX NCBI_TaxID=1587;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=53/7;
 RA Smeds A., Puteri T., Palva A.;
 RT "Identification of a gene cluster for the mevalonate pathway in
 RT Lactobacillus helveticus.";
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBD databases.
 DR EMBL; AJ279018; CAC51372.1; -.
 DR InterPro; IPR006204; GHMP_kinase.
 DR InterPro; IPR006206; MEV_galkinase.
 DR InterPro; IPR005917; Pmev_kin_Gpos.
 DR Pfam; PF00288; GHMP_kinases; 1.
 DR PRINTS; PR00959; MEVGALKINASE.
 DR TIGRFAMs; TIGR01220; Pmev_kin_Gr_pos; 1.
 KW kinase.

SO SEQUENCE 294 AA; 32537 MW; 0795099A1BC97938 CRC64;

Query Match 6.3%; Score 141.5; DB 2; Length 294;
 Best Local Similarity 22.2%; Pred. No. 0.029;
 Matches 75; Conservative 59; Mismatches 115; Indels 89; Gaps 19;

QY 77 VQSRINPLEATIFIVLAVIOPTFAE---DLEIITISDPGVHSQEDETKTSNGEKT 132
 DB 3 IDNRDNP-----EYLISATSYEQYCEIQNKKMYD---LHNSDLD---SADG----- 47
 QY 133 LYHSRAITEVEKTKGSSAGLVSVVATSLSHFIPNVISTNKDILHNAVAQIAHCAQAQKI 192
 DB 48 -----KKYGGSSNA--VTVATVAKALIFYG--VKMSNELYKKSALISH--YSVQGN 93
 QY 193 GSGFVATAIV--GLIVRRF-----QPALINDVFOVLESDEPKFTELKLESNME--X 245
 DB 94 GSAGDILASVYGWLAIVOTFPDKKLQYELANKT-----LSDVNEAMPGLK 139
 QY 246 HERCTLPYGIKLMGDVKGSETPKIVSRVLOWKEKEESSVVDOLNSANLQFMKEIR 305
 DB 140 IELTPPHDMKLMIG-----W--SQKASTSRIVDETNNANKAALNTEYK 181
 QY 306 EMREKTYSDPEYIKEL--DHSVEPLTVAIKIRK--LQALTQSEVPIEPDVOTQLD 361
 DB 182 NF---LAASSSECVLMKIVGFKAKNIALIKQIRVNRHLQHPAKINQAIETIRLTOLIK 238
 QY 362 RCOEIRPGCVGVVPGAGYDAIVL-----VLENO 391
 DB 239 IAEPRG--AAKTSAGNGDCGIVTTDADTDVDALENE 274

RESULT 18

O51630 PRELIMINARY; PRT; 317 AA.

AC O51630;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE Phosphomevalonate kinase, putative.
 GN BB0687.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=96065943; PubMed=9403685;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA Van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Uterback T., Watthey L., McDonald L., Artach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia
 RT burgdorferi.";
 RL Nature 390:580-586 (1997).
 DR EMBL; AE001169; AAC67030.1; -.
 DR TIGR; BB0687; -.
 KW Complete proteome.

SO SEQUENCE 317 AA; 35259 MW; CADB7E15350F88AB CRC64;

Query Match 5.8%; Score 129; DB 16; Length 317;
 Best Local Similarity 22.1%; Pred. No. 0.26;
 Matches 93; Conservative 52; Mismatches 126; Indels 150; Gaps 22;

QY 4 AFSAPGKAFLAGYVLEPIYDAYVATLSSRMHVAITPKGSLKSRKISSPOF-----A 59
 DB 5 SFSVPGNLLMGBETILE-----EKIGL---AIAIKRAFPSPFK 42
 QY 60 NGEWEYHISNTEKPRE---VQSRINPLEATIFIVLAVIOPTFADEIITISDPGVHS 116

Db 43 SDSWRF--SKKKKIDPFSLIENRSD-----FVKKAFVLSQNCFNLENFAY----- 88
 Qy 117 QBDTEKTSNGEKTFLYHSRATETVEKTLGSSAGL--VSVAATLSLHPHIVISTND 175
 Db 89 --DVIYDTSN-----PFFN-----DGTCKGPGSSAVVAIGVGLFLHNATNVV--EKG 134
 Qy 176 ILHNVAQIAHCVAKKKIGSGFVVAIAIYGLIV-----YRFPALINDVFOVLES 226
 Db 135 ELFKYGLAEAVRSQGGIGSGYDIATSI PGGVLEFEGFNPCKQQLGAVENDEPLYMGQGL 194
 Qy 227 PEKPELKLIESNNEEKEKCTLPYGIKLMGVKGSSETPKLYSRVLQMKKEKPESS 286
 Db 195 AIKTTT---SICEYN---KRRNSIDPILKCNL-----EMKKLV----- 227
 Qy 287 SVAYDQANSANLQ--FMKELREMRKCYSDPEYIKELDHSVEPLTVAIKIRKGLALT 344
 Db 228 -----LANSNKSALISLRKKE-----LGALG-----EALG 256
 Qy 345 QKSEVPIEPDVOTQLDRCEI PCGVGVVPGAG-----GYDAI AVLLEN 390
 Db 257 VSAALPSSFD--HLLGQCDLIAL-----GAGNETPLVYRPNIEAFNLKISIVLEN 307
 Qy 391 Q 391
 Db 308 E 308

RESULT 19

Q9STBI PRELIMINARY; PRT; 376 AA.

AC Q9STBI; 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Mevalonate kinase (EC 2.7.1.36).
 GN MK.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.
 OC NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia.
 RA Lluich A., Arro M., Riou C., Karst F., Boronat A., Ferrer A.;
 RT "Cloning and sequence of the gene encoding mevalonate kinase from
 Arabidopsis thaliana.";
 RL EMBL, L77688; AAD45421.1; -
 DR Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; L77688; AAD45421.1; -
 DR InterPro; IPR006203; GHMPkinase ATP.
 DR InterPro; IPR006204; GHMP kinase.
 DR InterPro; IPR006206; Mev_galkinase.
 DR InterPro; IPR006205; Mev_galkin.
 DR Pfam; PF00288; GHMP_kinases; 1.
 DR PRINTS; PR00959; MEVGALKINASE.
 DR TIGRPFAM; TIGR00549; mevalon_kin; 1.
 DR PROSITE; PS00627; GHMP_KINASES_ATP; 1.
 KW kinase, transferase.
 SQ SEQUENCE 376 AA; 40443 MW; E61B6E7E86D9474 CRC64;

Query Match 5.7%; Score 127; DB 10; Length 376;

Best local similarity 20.0%; Pred. No. 0.47;

Matches 88; Conservative 72; Mismatches 143; Indels 136; Gaps 17;

Qy 7 ARPKAFPLAGVLEPIYDAVYATLSSRMHAYIT-----PKGTSLKESRIKISSPOFANG 61
 Db 7 ARGKILAGEHAV---VHGSTAVAAIDYTYVTLRFPPLPSNDRLLQLKDISLFFS-- 61
 Qy 62 EWEYHSHSTKREKREVSQRIINPFLAETIFIVAVIQPTAEDELIITISDPGYHSGE--- 118
 Db 62 -MS-----LARIKALPYDSSTICRSTIPASCSSETLK 92
 Qy 119 -----DTEKTSNGEKTFLY-HSRAL-----TEVE-KTGLSSAGLVS 155

Db 93 SLAVLEBQNPKEKMWLSSGISTFLMYLTRYIIGNPATVIVNSLELPYSSGSSAALCV 152
 Qy 156 VATLSLSHFT-----PVIISTNDILHNVAQIAHCVAKKI-----GSGFVVAIA 202
 Db 153 ALTAALLASSISEKTRGNGWSSLDFTNLELKNW-----FEGEKI HKGKPGINDYVA 207
 Qy 203 YGLIYRRPQALINDVFOVLESDEKPETEKLIESNNEEKEKCTLPYGIKLMGV 262
 Db 208 YGNMI-----KFGSGEITLQSN-----MP--LRMLITWT 235
 Qy 263 KQGSSETPKLYSRVLQMKKEKPESSVYIDQANSANLQFMKELREMRKCYSDPEYIKEL 322
 Db 236 RVGRNTYKALVSGVSOAVRHPDMKSVFNAVDISKEKLAIIQSKDETSVTEKEKIEL 295
 Qy 323 DHSVEPLTVAINIKIRKGLALTQKSEVPIEPDVOTQLDRCEI PCGVGVVPGAGYDA 382
 Db 296 MEMNGGLLSMGVSHSSIEAV-----LITTVKHLVSK-----LTGAGGGGC 337
 Qy 383 IAVLVLENQVGNFKOKTLE 401
 Db 338 VLTLL---PTGVVDKQVIE 353

RESULT 20

Q63714 PRELIMINARY; PRT; 3187 AA.

AC Q63714; 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE GIANTIN (Golgi complex-associated protein of 364 kDa) (GCP364).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=KIDNEY;
 RX MEDLINE=98093490; PubMed=94311462;
 RA Toki C., Misumi Y., Fujiwara T., Sonda M., Nishioaka M., Ikehara Y.;
 RT "Identification and characterization of rat 364-kDa Golgi-associated
 RT protein recognized by autoantibodies from a patient with rheumatoid
 RT arthritis.";
 RL Cell Struct. Funct. 22:565-577(1997).
 CC -!- FUNCTION: MAY BE INVOLVED IN THE FORMATION AND/OR MAINTENANCE OF
 CC THE CHARACTERISTIC GOLGI STRUCTURE.
 CC -!- SUBUNIT: DISULFIDE-LINKED HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI MEMBRANE.
 DR EMBL; D25543; BAA05026.1; -
 KW Golgi stack; Antigen; Coiled coil; Transmembrane.
 FT DOMAIN 1 3163 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 3164 3184 POTENTIAL.
 FT DOMAIN 3185 3187 LUMENAL (POTENTIAL).
 FT DOMAIN 47 116 COILED COIL (POTENTIAL).
 FT DOMAIN 126 398 COILED COIL (POTENTIAL).
 FT DOMAIN 418 479 COILED COIL (POTENTIAL).
 FT DOMAIN 505 544 COILED COIL (POTENTIAL).
 FT DOMAIN 632 891 COILED COIL (POTENTIAL).
 FT DOMAIN 924 993 COILED COIL (POTENTIAL).
 FT DOMAIN 1010 1076 COILED COIL (POTENTIAL).
 FT DOMAIN 1103 1178 COILED COIL (POTENTIAL).
 FT DOMAIN 1231 1327 COILED COIL (POTENTIAL).
 FT DOMAIN 1348 1712 COILED COIL (POTENTIAL).
 FT DOMAIN 1765 2320 COILED COIL (POTENTIAL).
 FT DOMAIN 2331 2706 COILED COIL (POTENTIAL).
 FT DOMAIN 2356 2359 POLY-GLU.
 FT DOMAIN 2723 2775 COILED COIL (POTENTIAL).
 FT DOMAIN 2800 2892 COILED COIL (POTENTIAL).
 FT DOMAIN 2953 3008 COILED COIL (POTENTIAL).
 FT DOMAIN 3069 3113 COILED COIL (POTENTIAL).
 SQ SEQUENCE 3187 AA; 364295 MW; 48F035DFA3647F51 CRC64;

```

Query Match      5.5%; Score 122.5; DB 11; Length 3187;
Best Local Similarity 17.5%; Pred. No. 20;
Matches 80; Conservative 70; Mismatches 155; Indels 153; Gaps 13;

QY 28 VTALSSRMNAVITPDKTSLKESRIKISSPOFANGEMHYHSSNTEKPREVOSKINPFLA 87
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1373 IKALHTOLEMOAHEERLKQOVEICEKQPKLEBESKAQOOLQKQALISRKKA 1432
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 88 TTFIVLAIQPTFAFDLEITIIYSDPGVH---SOEDTEKTS-SNGEKTFYHSRAITVE 143
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1433 -----LKENSLOBOLSSADVAHEHLTKLADVESQVSVOQEKDALIGKALLQSE 1484
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 144 KTGSSAGIVSVVATSLSHFIPNVISTNKDILHVAQIACYAOKKIGSGFDVATAY 203
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1485 RDKL-----IVEMDKSLLENQSGSCESLKLALGG----- 1515
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 204 GLIVRRFPALINDVFOYLESDEKFPTELK-----KLIES-NMEKHEKCTLPGITL 257
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1516 -----LTEDKELMKLEBESVRCSTKAESTWQEKHELOKEY--EV 1554
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 258 LMGDVKGSETPKLVSRVLQWKKEPDESSVYDQNSANTL-----QPMKEL 304
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1555 LLAGSVENVAERLOHVVESVROEKQF---YAKKLSAESDKREKQLQDAEQMEM 1611
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 305 REMREKYSDPEPTYKELD-----HSAVEPLTVAI 333
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1612 KEKMRKFAKSKQOKKILEEENDRLRAEQPVGANESMALLSSNAKKEELERITLEY 1671
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 334 KNIRKGLQALTQKSEVPFIPDVOTQLDRCQETPGCVGVVPRAGGYDAIIVLVENQV 393
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1672 KTLSEKFEALMAKNTLSEETNKLQVBAQEL----- 1704
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 394 NFKQKTLNDYFNHYVWDLSEQTEGVLEKPEDEYIG 431
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1705 --KQASLEF-----TEKSDPEVDVIEVTEAVNG 1731
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 21
Q9PJ75 PRELIMINARY; PRT; 449 AA.
AC Q9PJ75:
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein Cj0036.
GN Cj0036.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagsis K., Karyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668 (2000).
DR EMBL; AL139074; CAB72529.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 449 AA; 52712 MW; AFB11PFAE73DA8E CRC64;

Query Match      5.4%; Score 120; DB 16; Length 449;
Best Local Similarity 22.0%; Pred. No. 1.9;
Matches 64; Conservative 45; Mismatches 94; Indels 88; Gaps 12;

QY 193 GSGFVATATVGLIYRRRFPAL-----IND----- 218
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 16 GSFIDISTALTAQIIDKAKQEWLKKKEPDEEVNAKRAEYTKALNDLTQKIEQKLINE 75
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

QY 219 -VFOYLESDEKFPTEL-----KLIENMEKHEKCTLPGIT 255
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 76 QVSOQLLEKQFQEDQLQKONFQKBEKKNKHEMNMKIMQBLEKSKSELSEFLSI 135
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 256 KL-----LMGDVKGSETPKLVSRVLQWKKEPDESSVYDQNSANTLQPMKELREMRKY 311
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 136 KAENRRLKRBQKENERBERLKFQAKEAFKFEKQESKNLFEERKORLEFQKSTQEDDLKY 195
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 312 DSDPEPTYKELDHSVEPLTVAIKNIRKGLQALT-QKSEVPFIPDVOTQLDRCQETPGC 369
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 196 -KELETFNSVQAQKLED--AQRRIEQSQQLQGAABELLIEYIQSEYLSDEVKVP-- 249
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 370 VGVVPRAGGYDAIIVLVENQVNF-----KQKTLNDYFNHYVWVD 413
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 250 -----KGVNADCLH--IVKNFQNIQSLYESKTYE-----FKKELND 288
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 22
Q9CKM9 PRELIMINARY; PRT; 640 AA.
ID Q9CKM9:
AC Q9CKM9:
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein PM1582.
GN PM1582.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pw70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pw70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
DR EMBL; AE006195; AAK03666.1;
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_transporter; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 640 AA; 72118 MW; 7A1D899F08820AF6 CRC64;

Query Match      5.3%; Score 118; DB 16; Length 640;
Best Local Similarity 21.6%; Pred. No. 4.4;
Matches 91; Conservative 69; Mismatches 146; Indels 116; Gaps 19;

QY 69 SNTKEPREVQSRINPFLFATTFIVLAIYQTEAFDL-EIITVSDP-----GYHSQE 118
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 268 AKASKAKQAOQRVYA-LERMELIAPAYVDNPFHFPREPLSLPMLAMQVSAQYPSAE 326
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 119 DTEKTSNGKRTFLYHSRAITVEYKGLSSAGIVSVATSL-----LSH 164
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 327 GQSAVEIIEKTKLNVPSRIGLKGNGAGKST-LIKLAGELTPLSGOVALANGVQIGY 385
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 165 FIPNVIST--NKDILHVAQIAHCYAOKKIG-----SGFDVATA 201
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 386 FAHQQLFRLRDESPFMHLOKLAQHHTQGLRDVLGPAFGDKVNEAVSFSGEKRL 445
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 446 VIALIIVQRPVILLDEPTNLLDMDRQALTEALVDYQSLVVSVDHRLRLNTVEBFYL 505
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 245 KHEKCTLPYGKIKLMGVKGSSETPKLVSRVLQWKKEPDESSVYDQNSANTLQPMKEL 304
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 506 VHDKKVEPFS-----GDLE--DYQKWLTVDNRTTEKADTNAVDANSSAN--RKE 553
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 305 REMREKYSDPEPTYKELDHSVEPLTVAIKNIRKGLQALTQKSEVPFIPDVOTQLDRL 362
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Db 554 OKRE-----AELRQQTAPL-----RKKLQOLEQKMEVSHLGOLEAQLVBS 596
 QY 363 COEIPGCVGVVPGAGYDAIAVLVLNENYGNFKO-KTLENPFYHNVYVWLEBETEV 421
 Db 597 DLYL-----AENKEKLTAL-LNEOVSNKCHLETLE-----MDMLAQEELEGL 638
 QY 422 LE 423
 Db 639 LE 640

RESULT 23

Q9FIM1 PRELIMINARY; PRT; 1088 AA.
 AC Q9FIM1;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosid 1; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=9915623; PubMed=10048488;
 RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
 RT Sequence features of the regions of 1,081,958 bp covered by seventeen
 RT physically assigned p1 and TAC clones.";
 RL DNA Res. 5:379-391(1998).
 DR EMBL; AB016885; BAB09633.1;
 SQ SEQUENCE 1088 AA; 122745 MW; 45A63120A97EF3C6 CRC64;

Query Match 5.3%; Score 117.5; DB 10; Length 1088;
 Best Local Similarity 23.1%; Pred. No. 10;
 Matches 113; Conservative 67; Mismatches 180; Indels 129; Gaps 27;

QY 40 TPKGTSLKESRIKISSPOFANGEME--YH-----ISSNTEKPREVOSRI--NPLEATI 89
 Db 608 TBSSES-----DVSKEPT--SSGRFEGMLPHTSASLSITEBPTLIDSIDGVNSEIMNSL 660
 QY 90 FIVLAVIOTPEAFDLEI-----IYSPGYHSGEDTETKTS 125
 Db 661 TSELTDQRPRLTSLSEMNLIDEVADMOQIENDLCSGPKIIDPDIIDHOQTD-QTSDS 719
 QY 126 SNGE-----KTFLYHS-----RAITEVEKTIG-----SSAGLVSV 156
 Db 720 IQGEHEHTSFLDASLDTPPIESFEREVOEBESNDKSTEETKTESBKLSSPQGV- 778
 QY 157 VATSLLSHPI-----PNVISNKKILHNVVAOIAHCYAKKI-----GSGFV 198
 Db 779 --TELLESVYVRENGEQLVKSADKAMLYEKEKTHNVLEASSSSNATQVLDYGAENS 836
 QY 199 ARAIYLYIYRRQPLINDVFOVLESDEKPFTELKLI--ESNNEKH-----ERC-TL 251
 Db 837 SVVIL-LQVDSHKSLPDESVDQELSKVEK--TELLKDFCGESTQTEYKRGVNEACGNA 893
 QY 252 PGYIKLMDGVKGSSETPKLVSRVLQMKKEKPESSVVYDQNSANLQFMKELREMR- 309
 Db 894 ENASDVLTLQVQGNNSPLDESTDQELSKVE--EKEVTLKDFDETPQGYKRRANVEES 951
 QY 310 -----KYSDPEYTIKE--LDHS--VEPLVIAIKIRKQLQALTQKSEVPLEPDVOTL 359
 Db 952 VVLADQNQODQGTWQOCCGIDSSGSIPTLEI-----TQLEDDVDVYASSSNVSG- 1004
 QY 360 LPRCOEIPGCVGVVPGAGYDAIAVLVLNENYGNFKOTLE--NDYFHHNVYVWLE- 415
 Db 1005 -DINKDITAAVANDILAASBRNDEV-ILKIDQGVGEAMEKELFELNLSLHHNIGLVGERBD 1062

QY 416 BQTEGVLES 424
 Db 1063 BESKVIIE 1071

RESULT 24

Q97SH9 PRELIMINARY; PRT; 335 AA.
 AC Q97SH9;
 DT 01-OCT-2001 (Tremblrel. 18, Created)
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
 DE 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 OS Phosphomethyltransferase kinase.
 GN SP0383.
 NCBI_TaxID=1313;
 RP Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 NCBI_TaxID=1313;
 (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouli H., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter C.M.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Frazer C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae.";
 RL Science 293:498-506(2001).
 DR EMBL; AB007350; AAK74550.1;
 DR TIGR; SP0383;
 DR InterPro; IPR005917; Pmev_kin_Gpos.
 DR TIGRfam; TIGR01220; Pmev_kin_Gr_pos; 1.
 KM Kinase; Complete proteome.
 SQ SEQUENCE 335 AA; 37072 MW; 6E415D098324392C CRC64;

Query Match 5.2%; Score 117; DB 16; Length 335;
 Best Local Similarity 22.4%; Pred. No. 2.1;
 Matches 95; Conservative 53; Mismatches 146; Indels 130; Gaps 20;

QY 9 GKAFIAGVIVLEPIYDAVYVVALSSRMAVITPKGTSLKES-RKISSPOFANGEMEYHI 67
 Db 8 GRLVWAGETALILEPQQLALINDIPIYKRAEI-----AFSDSYRYSMDPFA----- 54
 QY 68 SSNTEKPREVOSRI--NPLEATIYLAIV-----QTEAFDLEIYISDPGYHSGEDT 120
 Db 55 -----VLRPRPDVSLIOETALMGDPLAARGQMLRPFSLIC----- 92
 QY 121 ETKTSSNCKETFLYHSRAITEVEKTGLSSAGLVSVATSLSHPIPNVISTNKDILHNV 180
 Db 93 -----GKMEREGKFKGLSSGSVVVLVYKALLALY-----DVSQDLLEKFL 134
 QY 161 AQIAHCYAKKIGSGFDVATAI-YGLIYRRFPQPLINDVFOVLESDEKPFTELKLI 239
 Db 135 TS-AVLLRKGNGSGVDLACTVAEDLVLYQSPDR--QVAAWLEEE-----NATVLE 184
 QY 240 SNWEEKHERCTLPGYIKLMDGVKGSSETPKLVSRVLQMKKEKPESSVVYDQNSANLQ 299
 Db 165 RDM-----GFSI--SQVKTLECDPLVG-----WTEEVAVSSHMVOQIKONTION 227
 QY 300 FMKELREMRKXSDPEYTIKELDHSVEPLVIAIKIRKQ-LQALTQKSEVP-----IE 352
 Db 228 F-----LTSKKEVYSLVEALEQGSSEKIIDQVEVASKLLEGL 266
 QY 333 PDVQQLDRCOEIPGCVGVV--PGAGYDAIAVLVLNENYGNFKOTLENDYFHHNVY 410
 Db 267 TDIVYPLRLORKEASQDLQTVAKSSGAGGSGIALSPDAQ-----STYTLKN-----R 315

QY 411 WVDL 414
DB 316 MADL 319

RESULT 25

Q8DR49 PRELIMINARY; PRT; 335 AA.
AC Q8DR49; (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
GN Phosphomevalonate kinase.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=171101;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Albom W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA Dehoff B.S., Estrem S.T., Filtz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McArthur S.M., McHenry M., Mcleaster K., Mundy C.W., Nicase T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Kosteck P.R. Jr., Skatrud P.L.,
RA Glas J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RT J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008415; AAK9144.1; -
KW kinase: Complete proteome.
SQ SEQUENCE 335 AA; 37026 MW; B37D78CE13EED02E CRC64;

Query Match

Best Local Similarity 5.2%; Score 117; DB 16; Length 335;
Pred. No. 2.1;

Matches 92; Conservative 54; Mismatches 147; Indels 132; Gaps 20;

QY 9 GAAFLAGGLVLEPIYDAVYVTLSSRNHNVITPKGSLKESRIKISSPOFANGEWYHS 68
DB 8 GGLYAGVFAILEPQQLALIKDIPYMRALIA-----FSD---SYRIV 47
QY 69 SMTKEPREVQSRINPFLBATEFIVLAIQTEAPFLEIIYSDPGYHSGEDTETSS-- 126
DB 48 S-----DMFPAVDLRPNPDYSLIQETIALLMGDFL 77
QY 127 --NGEK---TFLYHRAITVEKGTGLSSAGLVSVATSLSHFIPNVISTNKDILHNV 180
DB 78 AVRGQNLRFPSLAIVGKMEREGKFGSSGVVVLVYVALLALY---NLSDQNLPLK 134
QY 181 AQAICVAKOKTIGSGFDVA--TAIYGLIYRRFQPLINDVFOVLESDEKPEFTELKITE 239
DB 135 TS-AVLTKRGDNGSGDLACIAEDLVLYQSFDR---QKVAMLEEE-----NLATYLE 184
QY 240 SNWEKHEKRECTLPGIKILMGDYKGGSETPKLVSVLVLMKKKEKRESSVYVDQNSALQ 299
DB 165 RDM-----GFSI--SQVKPTLECDPLVG---WTKVAVASSHVVOIKONION 227
QY 300 FMKELREKREKYDSDPEYIK-----ELDSVEPLTVAIKIRKGLQALTKSEVPI 351
DB 228 FL-----TSKKEIVLSLVEALBQKSEKIIIEOVAVSK--ILBGLST----- 267
QY 352 EBDVOTQLDRQCEIFPGCVGVV--PGAGGDALAVLVLENQVGNFKQKTELENDFYFNV 409
DB 268 --DIYPLRLQLEKASODLQAVAKSGAGGDCGIALSFDQ---STTKLKN----- 314
QY 410 YVVDL 414
DB 315 RMADL 319

RESULT 26

Q8SUB3 PRELIMINARY; PRT; 352 AA.
AC Q8SUB3; (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
GN Hypothetical protein ECU10_1510.
OS Eucaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
NCBI_TaxID=6035;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RX MEDLINE=21576510; PubMed=11719806;
RA Katinke M.D., Duprat S., Cornilliot E., Metenier G., Thonarat F.,
RA Premsier G., Barbe V., Peyretailade E., Brotlier P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Encephalitozoon cuniculi.";
RL Nature 414:450-453(2001).
DR EMBL; AL590449; CAD25871.1; -
DR InterPro; IPR006204; GMP_kinase.
DR Pfam; PF00288; GMP_kinases; 1.
KW Hypothetical protein.
SQ SEQUENCE 352 AA; 39394 MW; 510AD2EB3AC467B3 CRC64;

Query Match

Best Local Similarity 5.2%; Score 116.5; DB 5; Length 352;
Pred. No. 2.4;

Matches 86; Conservative 61; Mismatches 133; Indels 147; Gaps 20;

QY 5 FSAGKAFGLGGVLEPIYDAVYVTLSSRNHNVITPKGSLKESRIKISSPOFANGEW 64
DB 7 FRVPGKIVNGSYIVLS---GETCRAVALKTHMTSEATRTVSSAEITVA--VDGSKI 60
QY 65 YHISNTEKPREVQSRINPFLBATEFIVLAIQTEAPFLEIIYSDPGYHSGEDTETSK 123
DB 61 TYGSDYIRTCNDTCGPAHYILINIVACFEIARISPKRINHE--MHTGEGFSS----- 111
QY 124 TSSNGEKTFLYHRAITVEKGTGLSSAG--LVSVVATSLSH----- 164
DB 112 GSPG-----EKTGIGSSACILVSIYVALLRFHQDPFRIRASRGDPDAGR 157
QY 165 ---FIPNV-----ISTNKDILHNVAAQIANC--YAQKIGSGPVATATVGLIYRRFP 213
DB 158 SPVQSGDLKMLGSLISEDAEHLISITYLVHQAQNGASGCVWCCLGSIYFSR-- 214
QY 214 ALINDVFOVLESDEKPEFTELKILBSNWEKHEKRC---TLPGIKILMGDYKGGSETP 269
DB 215 -----ERCFPLENIRP--YLIIGSPKSSATR 239
QY 270 KLVSRVLMKKKEKRESSVYVDQNSAN--LQFMKEL--REKREKYDSDPEYIKELDSV 326
DB 240 EMLKKI-----NSKDPKWRPFKDINRKIRNEBESPKLKYMEYLD--- 278
QY 327 EPLTVAIKIRKGLQALTKSEVIEBDVOTQLDRQ--EIPGCVGVVPGAGGYAIA 384
DB 279 ---ANKGV-----STAIVEKQYEVLMKNGYDIWC---GISGAGGDCVW 319
QY 385 VLVLNQ 391
DB 320 ALADDYQ 326

RESULT 27

Q43191

ID Q43191 PRELIMINARY; PRT; 862 AA.
 AC Q43191;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
 DT 01-UN-2002 (Tremblrel. 21, last annotation update)
 DE Lipoxigenase (EC 1.13.11.12).
 GN POTLX-3.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamiales; Solanales; Solanaceae; Solanum.
 OC NCB1_TaxID=4113;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Beroлина;
 RA Kolomoets M.V., Hannapel D.J., Gladon R.J.;
 RT "Nucleotide Sequence of a cDNA Clone for a Lipoxigenase from Abscissic
 RT Acid-Treated Potato Leaves (Accession No. U60202) (PGR96-069).";
 RL Plant Physiol. 112:446-446(1996).
 DR EMBL; U60202; AAB67865.1; -.
 DR HSSP; P08170; 2SHL.
 DR InterPro; IPR000907; Lipoxigenase.
 DR InterPro; IPR001024; Lipoxigenase_LH2.
 DR Pfam; PF00305; Lipoxigenase; 1.
 DR Pfam; PF01477; PLAT; 1.
 DR PRINTS; PR00087; LIPOXYGENASE.
 DR SMART; SM00308; LH2; 1.
 DR PROSITE; PS00711; LIPOXYGENASE_1; 1.
 DR PROSITE; PS00081; LIPOXYGENASE_2; 1.
 KW Oxidoreductase.
 SQ SEQUENCE 862 AA; 97778 MW; 94667A9F56041E89 CRC64;

Query Match 5.2%; Score 116; DB 10; Length 862;
 Best Local Similarity 18.4%; Pred. No. 9.4; 115; Indels 166; Gaps 17;
 Matches 76; Conservative 56; Mismatches 115; Indels 166; Gaps 17;
 24 YDAVYATLSSRMHAYITPKGTSLK-----ESRIKISSPOFA--NG 61
 302 PDSFEFVLMKYEGGKILPGQPFILKALDLSIPLEIKRIITDGEKFKPTPQVIGEDS 361
 62 EWEYHISNTEKREVSQRINPLEATITFVLAYIQ--PTFAFDLEIITISDPCYHSE 118
 362 SWR---TDEEFAREMLAGVNP-----VILSRLOEFPKSQLDSEV--YGNQN 403
 119 DRETK-----TSSNGEETPLYSRAITREVK 144
 404 STTKHEHIENTLDGLITIDAIKTNRLYLINHDILMPYVRRIITNFKLPASTKLPLDQ 463
 145 TGL-----GSSAGLVSVVAT-----SLASH 164
 464 DGTMKFVAIELSLPHPDGDELGAIVSKYTPADQGVSGSIWQLAKAVAVANDSGVHQLISH 523
 165 FI-----PNVISTNK--DILHNVAQIAHCYAKKIGSGFDVAIYGLIVRRFQPL 215
 524 WLNTHAIEFPVATINROLSVLPHIKLHPFRDITN-----INALARQL 570
 216 INDVFOVLSD--PEKFPTELKLIISNNEEKH-----ERCTLPYGIKILM 259
 571 IN-AGVLENTVTPAKYAMEMSAVYKSWFPEQALPADLIKRGAVVEDSSSPHGRLLI 629
 260 GDVKGSETPKLVSRVLQWKE-----KPESSVYVDOLNSANLOFMKELR 306
 630 QDYPAVADGELISAKSWTEVCNRYYSDELVLQNDLEQA---MWKELR 678

RESULT 28
 Q9FKNS PRELIMINARY; PRT; 1342 AA.
 AC Q9FKNS;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, last annotation update)
 DE Gb|AAC35233.1|.

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosida II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCB1_TaxID=3702;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=98344145; PubMed=9679202;
 RA Kaneko T., Kocani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence
 RT features of the regions of 1,381,565 bp covered by twenty one
 RT physically assigned P1 and TAC clones";
 RL DNA Res. 5:131-145(1998).
 DR EMBL; AB011480; BAB11225.1; -.
 SQ SEQUENCE 1342 AA; 149599 MW; 5A91BAD0A21855F2 CRC64;

Query Match 5.1%; Score 114; DB 10; Length 1342;
 Best Local Similarity 22.0%; Pred. No. 24;
 Matches 92; Conservative 41; Mismatches 135; Indels 150; Gaps 21;
 21 EPIYAYATLSSRMHAYITPKGT-----SLKSRKISSPOFANGEWYHISNTEKPR 75
 748 EBYVDS-----SPSGSRFPFSVSSVDYKPDLE-KNGE--EIEENEKER 791
 76 EYQSRINPLEATITFVLAYIQPTFAFDLEIITISDPCYHSEDETTSNGEKTPLYH 135
 722 EYVSE-----STGPEI-----HSTNTEYTSVGENSGMYT 825
 136 SRA--ITVEKTKLGSSAGLVSVVATSLSHRIPNVISTNDILHNVAQIAHCYAKKIG 193
 826 GSAIWMHEHSTPLESPDVVHDAET-----SVNKSVAEII--MYEEBAQ 870
 194 SGFD-----VATAYGLIVRRFQPLINDVFOVLESPEKFPTELKLI 238
 871 KOKDVSPTQFNADIPIDSVASLSGAVEYETHSFNDEDAQ--LEQEP-----VHSLV 923
 239 ESNMEKEKRCETLPYGIKILMGDVKG--GSE--TPKLVSRLQWK--KEKPESS----- 286
 924 HDAEBETHNDQTDIEVSVNASQNVGSEBTSPESSRELTWSDPKSVYEOSSLEPGDQ 983
 287 -----SVYVDOLNSANLOFMK-----ELREKREKYSDP 315
 984 VPTRAGPVSVF---SNITFHEYHDAPEDTTELSCLTSPTSSPTSPPTPMWEG 1039
 316 -----ETYEKLDHSVEPLTVAIKIRKGLQALTQKSEVPIEPDVOTOLLDRCOEI 366
 1040 SRAEFPQDIYEBLDVVERLEQL-----TDLHAISSQ--PEIITEADEIKEI 1087

RESULT 29
 Q9PD52 PRELIMINARY; PRT; 336 AA.
 AC Q9PD52;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, last annotation update)
 DE Phosphomevalonate Kinase.
 GN MVAK2.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OC NCB1_TaxID=1313;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20353468; PubMed=10894743;
 RA Wilding B.I., Brown J.R., Bryant A.P., Chalke A.F., Holmes D.J.,
 RA Ingraham K.A., Jordanescu S., So C.Y., Rosenberg M., Gwynn M.N.;
 RT "Identification, evolution and essentiality of the mevalonate pathway
 RT for isopentenyl diphosphate biosynthesis in gram-positive cocci";
 RL J. Bacteriol. 182:4319-4327(2000).
 DR EMBL; AF290099; AAG02457.1; -.

DR InterPro; IPR005917; Pmev_kin_Spos.
DR TIGRPFAM; TIGR01220; Pmev_kin_Gr_pos; 1.
KW Kinas.
SQ SEQUENCE 336 AA; 37167 MW; A2409E26569D2511 CRC64;

Query Match 5.1%; Score 113.5; DB 2; Length 336;
Best Local Similarity 21.8%; Pred. No. 3.7;
Matches 94; Conservative 58; Mismatches 136; Indels 143; Gaps 20;

QY 9 GRAFLAGYLVLEPIIDAVYVTAALSSRMHVVITPKGSLKES-RIKSSPQPMNGEMEHYI 67
DB 8 GLTVAGAEVAILLEPGQALIKDIPYKAEI-----AFSDSTRITSMDMFA----- 54
QY 68 SSNTEKPREVOSRINP---FLSATPIVLAIV---QPTFAEDLEIITVSDPGHSQEDPT 120
DB 55 -----VDLRPNPDYSLIQETIATMGDFLAVNGQNRPPSLKIC----- 92
QY 121 ERTKTSNGEKTFVLSRAITEVEKTLGSSAGLVSVATSLSHPIPNVISTNKDILHNV 180
DB 93 -----GKMEREGKKEFGLSGSGSVVVLVVKALLALY---NLSDQNLFLYL 134
QY 181 AQIAHCYAOKKIGSGFDVATAI-YGLIVYRRF--QPALI-----NDVFGVLESDEPKPTPE 233
DB 135 TS-AVLKRGDNGSMGDLACTVABEDLVLYQSPDRQAAALEENLAVLERDWGFFLSQ 193
QY 234 LKKLIESNMEEKHERCTLPYGIKILMGDVKGSGSETPKLVSRVLQWKEKEPSESSVYVDL 293
DB 194 VKPTLECDP-----LVG-----WTKVAVASNHAVQDIT 221
QY 294 NSANIQPMKELREMEKIDSDPEYIK-----ELDSVEPLTVAIKNIRKGLQALITQ 345
DB 222 QMINQNFL-----SSSKETVSVLVEALEQKAEKIVQVEVASK-LLEGIST--- 267
QY 346 KSEVPIEPVOTQLDRCEIEICGCVGVY--PGAGYVAILVLENGVGNKQKTLNLP 403
DB 268 -----DIYTPLRQLKEASQDLOAVKSSGAGGDCIALSPDAQ---SSRNTLKN- 315
QY 404 DYFNHYVVDL 414
DB 316 -----RMADL 320

RESULT 30

Q8KOK5 PRELIMINARY; PRT; 340 AA.
AC 08KOK5;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Pneumococcal surface protein A (Fragment).
GN PSPA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCB1_TaxID=1313;
RN NCB1_TaxID=1313;
RP SEQUENCE FROM N.A.
RC STRAIN=St 435/96;
RX MEDLINE=22170754; PubMed=12183557;
RA Miyaji E.N., Ferreira D.M., Lopes A.P.Y., Brandileone M.C.C.,
Dias W.O., Leite L.C.C.;
RT "Analysis of Serum Cross-Reactivity and Cross-Protection Elicited by
RT Immunization with DNA Vaccines against Streptococcus pneumoniae
RT Expressing Pspa Fragments from Different Clades.";
RL Infect. Immun. 70:5086-5090(2002).
DR EMBL; AY082387; AAL92492.1; -;
DR InterPro; IPR000533; TROPOMYOSIN.
DR PRINTS; PRO0194; TROPOMYOSIN.
FT NON_TER 1 1
FT NON_TER 340 340
SQ SEQUENCE 340 AA; 38023 MW; EE07ECP00B1FBD57 CRC64;

Query Match 5.1%; Score 113; DB 2; Length 340;

Best Local Similarity 23.8%; Pred. No. 4.1;
Matches 59; Conservative 44; Mismatches 87; Indels 58; Gaps 12;

QY 210 RQPALINDVQVLESDPEKPTTELKLIENNMEEKHERCTLPYGIKILMGDVKGSGSTP 269
DB 82 KYQDELVY---YIRENDPTK-KAEKAKMDABEKYKKQT-----EFAEVRA----- 125
QY 270 KLVSRVLQWKK--EKPESSVYVDQLNSANIQPMKELREMEKIDSD-----PEYIKY 321
DB 126 KVIPSABELKKTQKAEBAKKAKEALTKVBEAEKQVTEAKQYDAEHAEEVAPQAKIAE 185
QY 322 LDHSVEPLTVAIKNI-----RKGLQALTQKSEVPIEPVOTQLDRCEIEICGCVG 372
DB 186 LEHEVQKLEKALKEIDESSESDYVEGIRAPLQ-----FELDVQAKLSLEFLSDKIDE 240
QY 373 VVPAGGDAVAVLVEENGVPKQKTLNENDYHNHYVNDL-----EQTGEVLE--- 423
DB 241 L-----DA-EIAKLEKVDVDFKNSDGEQGLAAAEEDLVAKKAELEKTEADLKAV 292
QY 424 ---EKPED 428
DB 293 NEPEKPAE 300

RESULT 31

017102 PRELIMINARY; PRT; 733 AA.
ID 017102;
AC 017102;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein F42G2.6.
GN F42G2.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitae; Rhabditoidea;
OC Rhabdilitae; Pelodierinae; Caenorhabditis.
OX NCB1_TaxID=6239;
RN NCB1_TaxID=6239;
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN NCB1_TaxID=6239;
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Scheet P., Magg L.;
RT "The sequence of C. elegans cosmid F42G2.";
RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
RN NCB1_TaxID=6239;
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (DEC-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF024499; AAB70358.3; -;
KM Hypothetical protein.
SQ SEQUENCE 733 AA; 85879 MW; 71913271DB8DFDE3 CRC64;

Query Match 5.0%; Score 111.5; DB 5; Length 733;
Best Local Similarity 30.0%; Pred. No. 16;
Matches 57; Conservative 24; Mismatches 56; Indels 53; Gaps 12;

QY 193 GSGFDVATVYGLIVRRFQPALINDVQVLESDEKPTTELKLIENNMEEKHERCTLP 252
DB 534 GSGGLTISEA-HGVY---GSPMLI-EKKELEGLIEHQPTMANEIDQNKKE--ETTKP 585
QY 253 -YGIKILMGDVKGSGSETPKLVSRVLQWKEKEPES-----SVYVDL----- 293
DB 586 EYRKQLMSRKQNEILLQKL-----WLKEKNTENTHAETKNOEKKLTENAIVOQLKQNM 639
QY 294 -----NSANIQPMKELREMEKIDSDPEYIKELDSVEPLTVAIKNIRKGLQAL 343

Db 640 KMIQERRAADSXNKLKEQVLEMKDKLDAQ-ETRIKELEEVK----SDKNAKGEQ-L 693
 QY 344 TOKSEVPIEP 353
 Db 694 TD---PTEP 699

RESULT 32

Q91423 PRELIMINARY; PRT; 642 AA.
 ID Q91423
 AC Q91423
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Serine/threonine kinase.
 GN PKNC.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PCC 7120;
 RA Gonzalez L., Phalip V., Zhang C.C.;
 RT "Phosphorylation of the signal transduction protein PII by the Ser/Thr
 RT kinase PncC in the cyanobacterium Anabaena sp. strain PCC 7120."
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A251821; CAB75357.1; -
 DR PhosSite: Q91423; -
 DR InterPro: IPR00719; Prot_kinase.
 DR Pfam: PF00069; Pkinase; 1.
 DR ProDom: PD000001; Prot_kinase; 1.
 DR PROSITE: PS50011; PROTEIN KINASE_DCM; 1.
 DR ATP-binding; Kinase; Transferase
 SQ SEQUENCE 642 AA; 70785 MW; 81B36952C6E9E8B7 CRC64;

Query Match 5.0%; Score 111; DB 2; Length 642;
 Best Local Similarity 23.7%; Pred. No. 14;
 Matches 101; Conservative 54; Mismatches 160; Indels 112; Gaps 24;

QY 25 DAYVT-----ALSSMHAIVITPKGSLKESRIKISSPQFANGEMVYHSSNTEKREVO 78
 Db 257 DRQRTASEALQALONALNPVAVSVTSARS--LKNS-----YQSANPSSVSQ 304
 QY 79 SRI-----NPF-----LEATIFVLA-----YIQTEAPDLRI-- 107
 Db 305 QIVAAVPTPVYTKPKGRKSHNSDPLPLIGIVLAGAALAVANFPVYKAPFANVLSNN 364
 QY 108 -----IYSDPGYHQEDTETKTSNGEKTFLYHSRAITEVEKTVLGSSA 151
 Db 365 ATSGNKCLAVVAGNSNIRSEF---SSINTDTVLQITGVN-----INFETVGRKTR 412
 QY 152 GLYSVAVATLSLHPIFNVIISTNKDIIHNVAQIAHCAQKKG-SGDVATAIYGLIYKR 210
 Db 413 GWEIKFNFS-----SRLMAHSDVITINDQWTSCTGKGIALKTYDSDTLAAREVPRP 466
 QY 211 POPAL---INDVQVLESDEPEKPELTKLIESNWE-----EKHERCTLPYGIKILMGD 261
 Db 467 -QPKLDSFINSAPER-EASFNQPTFAQPKEDTTVVAQARKKYSGLVGAIALMRBI 524
 QY 262 VKGSEETPLVSRVL-QMKKEPSESVVYDQNSA--NLOFMKELREMKREKYSDDPEY 318
 Db 525 PANASAGIKETSMINQWQDM-QKADALFNDINKALENGQMDKVL-----EYKQPEKL 578
 QY 319 --IKELDSVEPL-TVAIKIRKGLQALTKQSEVPIEP--DVOTQLDLRCQELPGCVGV 373
 Db 579 PINKWQDLEPVEFKQATEVNAK-QAL-PKTENPEPNNWMESESNVEESINETPSSE 635
 QY 374 VPQAGY 380
 Db 636 EPPQGY 642

RESULT 33

O23289 PRELIMINARY; PRT; 1249 AA.
 ID O23289

AC O23289
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical 135.8 kDa protein.
 GN A4G14310.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsidae.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9812113; PubMed=9461215;
 RA Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
 RA Bergkamp R., Dirksen S.A., van Staveren M., Stiekema W., Drost L.,
 RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
 RA Medler E., Wambutt R., Welzenegger T., Pohl T.M., Terry N.,
 RA Gietlen J., Villarreal R., De Clerck R., van Montagu M., Lecharny A.,
 RA Auborg S., Gy I., Kreis W., Lao N., Kavanagh T., Hempel S., Kotter P.,
 RA Eutlan K.D., Rieger W., Schaeffer M., Funk B., Mueller-Auer S.,
 RA Silvey M., James R., Montfort A., Pons A., Pulgomech P., Douka A.,
 RA Vouklatou E., Milioni D., Hatzopoulos P., Pizavandi E., Obermayer B.,
 RA Hilbert H., Duesterhoft A., Moore T., Jones J.D.G., Eneva T.,
 RA Palme K., Benes V., Reichman S., Ansoorge W., Cooke R., Berger C.,
 RA Delseny M., Voet M., Volckaert G., Mewes H.W., Klotterman S.,
 RA Scheller C., Chailwatiz N.;
 RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
 RT Arabidopsis thaliana."
 RL Nature 391:485-488(1998).
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z97335; CAB10210.1; -
 DR EMBL: AL161538; CAB78472.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 1249 AA; 135836 MW; 75B852EC34073CD4 CRC64;

Query Match 5.0%; Score 111; DB 10; Length 1249;
 Best Local Similarity 20.0%; Pred. No. 36;
 Matches 93; Conservative 70; Mismatches 159; Indels 142; Gaps 20;

QY 38 VITPKGSLKSRKISSPQFANGEMVYH--SSNTEKREVO--RIHPLEATIFVLA 94
 Db 167 VIGEGESGQDKSGKSGFGQGTSEIKVPSVCEKADGSSCPVNS----- 215
 QY 95 YIQTEAPDLRIIYSDPGYHQEDTETKTSNGEKTFLYHSRAITEVEKTVLGSSAGLV 154
 Db 216 --SKFEGSVARNSISDRAHA-----LVSGEKSTV-ALKSKSKIEKKGKTSVVL- 264
 QY 155 SVAVTSLSLHPIFNVIISTNKDIIHNVAQIAHCAQKKGSGDVATAIYGLIYKR 214
 Db 265 -----RRKSLDNVQKAMGMSKDRIQNESSNSTAKY-----PS 297
 QY 215 LINDVQVLESDEPEKPELTKL--LIESNWEK-----HERCTLPYGIKILMGYK 263
 Db 298 KLHEKLAFLGKVKVKIADIKTYKMDLNNPDSKVIISDIHOKIT---GIEKSMSHVI 354
 QY 264 GGSSE---TPKLVSRVLQMKKEPSESVVYDQNSA--NLOFMKELREMKREKYSDDPEY 318
 Db 355 DQPEKQKTKKAKSVKGLKSELEDRLLPHQRLRSRQSTSSHSVSGHDSVESNRAVN 414
 QY 294 -----NSANLQFMKELREMKREKYSDDPEY--IKELDSVEP----- 328
 Db 415 AEEKSAVEENVAIALBFLASLDEKEKVFEMDQNALENVQEMD-TREPKEKNDVSKDV 473
 QY 329 -LITVAIKIRKGLQALTKQSEVPIEPDVOTQLDLRCQELPGCVG-----GVPPAGGY-- 380
 Db 474 NLTSMILTEILRANALBEIIDEENREWELEIID---DQCMQQLANDIGKSTSTGWFV 528


```

QY 2 SKAFSAAPGKFLAGYLVEPIYDAYVATLSSRMHVAITPKGTSLESRIKISSPOFANG 61
D 10 SKNKSAPGIGFTAG-----STNGSGISIGL-VSSIG- NV 43
QY 62 EWEYHISNTEKP---REVQSRINPFLAETIFIVAYIOPTFAEDLEIIITYSDPGYHSGE 118
D 44 TSSQDISKRTNPKLMEKRRRRARINQSLALIKALIL-----E 79
QY 119 DRETKSSNGEKTFLHSAITEVEKGTGSSAGLVSVAATLSLHFIPIVISTNK----- 174
D 80 STKTQAKNGEGGAKH---TLERK-----ADILELTVRRHFORRNLDLDPVNRKTRAG 128
QY 175 --DILHNAQIARCAVQAKKIGSGFDVATAIYGLIVRRFOPALINDVFOLESDEPKFPT 232
D 129 YTDCAEVARILATPEPPMGT-----MPTL-----AEPGSKAR 162
QY 233 ELKKLIESWBEKHEKCTLPYGIKLMGVKSGSETPKLV--RVLOMKKEKPESSSVY 290
D 163 LIRHL-----DCCIAEIDVEICPHSTAAFAESPSSSCFDLHGKKSQPEERSLDY 213
QY 291 DQNSANLQFMKELRMRREKVDSPETIYIKEL-----DHSVEPLTVA--IKNIRKGLQAL 343
D 214 -----SSQDSNPVDYSKGLKMAVAEORTLPVTPAPQDENNNRGLQAO 255
QY 344 TOKSEVPIEDVOTQLDRQCEIPGCVGVPGAGGYDAIAVLLENQVGNFKOKTLENP 403
D 256 AQ-TPIPIVQSGTO---GQTSBH-VDAVABSELSEEDRNVCANVLEQYKQ----- 303
QY 404 DYFNHYWVLEBQTEGLEKPEDEYIGL 432
D 304 -----QLKAHVQOQOESANGV 319

RESULT 36
AC 029291 PRELIMINARY; PRT; 600 AA.
ID 029291;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein AF0971.
GN AF0971.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OC NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.U., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kierlavage A.R., Graham D.E., Kyriplides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirtress E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
EMBL; AE001037; AAB90272.1; -.
DR TIGR; AF0971; -.
DR InterPro; IPR000644; CBS domain.
DR InterPro; IPR000595; CNMP binding.
DR InterPro; IPR005105; DUF254.
DR Pfam; PF00571; CBS; 2.
DR Pfam; PF00027; CNMP binding; 1.
DR Pfam; PF03445; DUF294; 1.
DR SMART; SM00116; CBS; 2.

```

```

DR SMART; SM00100; CNMP; 1.
DR PROSITE; PS50042; CNMP BINDING 3; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 600 AA; 69122 MM; C74084D24F5229C1 CRC64;

Query Match
Best Local Similarity 18.7%; Pred. No. 14;
Matches 97; Conservative 77; Mismatches 146; Indels 199; Gaps 25;

QY 19 VLEPIYDAYVATLSSRMHVAITPKGTSLESRIKISSPOFANGEWEYHISNTEKP----- 74
D 136 ITRDLFLTRVGBELSKKVCVCSFY-TSIRDAIKRE-----LVGSGIYVVDNKLPLQIL 190
QY 75 -----REVQSRINPFLAETIFIVAYIOPTFAEDLEIIITYSDPGYHSGEDTET 122
D 191 TSKDPTFTIYKSGQEKVSAVMSPV-VAADYSTPVEAHLELL----- 234
QY 123 KTSNGEKTFLYHSAITEVEKGTGSSAGLVSVAATLSLHFIPI----- 167
D 235 -----KGINHLVVTENGKVRGVIT--ANDILTFEPTSLIYLVRKLKRAK 279
QY 168 ---NVISTNKDILHNAQIA---HCYAQKIGSGFDVATAIYGLIVRRFOPALINDV 220
D 280 SLEIKNTFKKLTYSISLVNRGMHFFDLS-----NLTETIYDLYVK----- 322
QY 221 OVLSEDPKFPTELKGLIESNW-----BEK-----HERCTLPGIKLMD 261
D 323 -VIEWEKRFERENGKLPSPAVHMGSSARKEQVATIQDNAIYHE-----GD 369
QY 262 VKGSGSETPKLVSRVL-----QMKKEKPESSSVV--YDQNSANLQFMKE 303
D 370 GEGLSQFASVNNALDYGVIPKRCGYMANMCKSVSEMKSVFSEMFIKLPDNLRLFLSV 429
QY 304 LEMAREKTDSPETIYIKEL-----DHSVEPL-----TVAIKNIRKGLQALTKSV 350
D 430 FLDLRVLVIGD--EKICKELIHKIKREHTSQSLRYLAYDATTAEPPL--GLFGLRKKKID 485
QY 351 IEPDVQQLDRQCEIPGCVGVPGAGGYDAIAVLLEN--QVGNFKOKTLENPDYFH 407
D 486 LK-----MNGIYIYNG---VRLALHEHLIETVNTRR--LEELKKAH 524
QY 408 NYWVWVLEB-----QTEGLEKPEDEYI 430
D 525 -VMDADLAEISLKETEFQLDRLRQAKEVVEGKKADNV 562

RESULT 37
AC 020949 PRELIMINARY; PRT; 790 AA.
ID 020949;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 88.6 kDa protein F59P5.7.
GN F59P5.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA White S.;
RT Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RA Leonard N.;
RL Submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL; Z50794; CAA90661.2; -.
DR EMBL; Z69646; CAA90661.2; JOINED.
DR EMBL; Z69646; CAA93477.2; -.
DR EMBL; Z50794; CAA93477.2; JOINED.
DR WormPep; F59P5.7; CE23791.
DR InterPro; IPR006609; DUF_DM13.

```

DR	SMART: SM00686; DM13; 1.
KR	Hypothetical protein
SO	SEQUENCE 790 AA; 88567 MW; A7C8E446BD634B9 CRC64;
Query Match	5.0%; Score 110.5; DB 5; Length 790;
Best Local Similarity	19.7%; Pred. No. 21;
Matches 101; Conservative 72; Mismatches 179; Indels 161; Gaps 24	
QY	9 GKAFIAGGLVLEPIYDAYVTALSSRMHAAVIT-----PKGTSLSKSR-IKISSPOFAN 60
DB	199 GVPMTAGGWETQPHYSG-----DRLHAASFVSGSPSIPATNVRKLRKIKNSLFGYN 253
QY	61 GEMEHISNTEKREPVOSRINPFLPATIFVLAAYIQPLEAPDL---EIIYY----- 109
DB	254 GE---LASALIQKQKVEPNITISOPATEKMLANEMSVESKSLANNLLGFNNGGTGE 309
QY	110 SDPGYHSEDDETETKTSNSGEKTFLYHSRAITEYEKTLGSSAGLVSVAATSLSHFIPNV 169
DB	310 SSGSGISLPEMTTSLVSGGNRK-VPTSSAAHDDQGYAPASAGRPQ---TSMLTQFNNKA 365
QY	170 ISTNKDILHNNAQIAHCAQOKKIGSGFDVAATAYGLIYRRPOPALLINDVFOYLES---- 225
DB	366 SDALKD-----SEFAKK-----VFENKENKND 387
QY	226 -----DPEK-----PPELTKLILSNWEKHKRCLTPYIKLIMGD 261
DB	388 PLALPIPEPEKEKEKESKEVENTLKLSTMPDEQKRLIESAIDGELIDADSPLIKNIYKND 447
QY	262 VKGSEETPEKLVSRVLQWKKK-----KPESSVYVDLQNSANLQFM-KELREME-KYDSQ 314
DB	448 VTSEKKER-ANRLLEMTIRSNRPSTKTPESGI-----KKNVYIKYKIGSLAETPEIAK 500
QY	315 PETYIK--ELDHSVEPLFVAIKNIRKGLQALQKSEVPI--PDVOTOLLDRQCE--IPG 368
DB	501 PRINVAAYKLFGPIPIYANVTRVXRKDSNVNATK-PIELFVAVSTWQINCKSNIGQ 559
QY	369 CVGCVVPAGAGIDALAVLVLENQVGNFKCKT-----LENPDY----- 405
DB	560 GIMEVRPANG-----VLKLEKQADVTYSMSVPGGLRLROPEFDPVDDKVPQPLQWFEQ 613
QY	406 -----FANVYVVDDEQTEGYLEEK 425
DB	614 FQPLMLTLTPRSKVFGDIHWLSLRDKRNILDYK 646
RESULT 38	
OS	08IJW2 PRELIMINARY; PRT: 3853 AA.
AC	08IJW2.
DT	01-MAR-2003 (TREMBLrel. 23, Created)
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE	Hypothetical protein.
GN	PF10_0079.
OS	Plasmodium falciparum (isolate 3D7).
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX	NCBI_taxid=36329;
RN	[1]
RC	SEQUENCE FROM N.A.
RP	STRAIN=3D7;
RX	MEDLINE=22255705; PubMed=12368864;
RA	Gardner M.-J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA	Carlton J.M., Pahn A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA	Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA	Chan M.-S., Nene V., Shalom S.J., Suh B., Peterson J., Angiolli S.,
RA	Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,
RA	McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA	Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA	Fraser C.M., Barrell B.;
RT	"Genome sequence of the human malaria parasite Plasmodium
RL	falciparum."
LT	Nature 419:496-511(2002).

DR EMBL; AE014830; AAN5277.1; -
SQ Hypothetical protein.
EMBL SEQUENCE 3853 AA; 453001 MW; 2BB8803E3195A92F CRC64;

Query Match 5.0%; Score 110.5; DB 5; Length 3853;
Best Local Similarity 18.5%; Pred. No. 2e+02;
Matches 74; Conservative 62; Mismatches 122; Indels 141; Gaps 17;

QY LEATFIYLAIOPTFAADLEIIIIYSDPG-----YHSQEDTETKT-SSN 127
DB LHSIYIPQSYDFDEVDQDIKTYEDPTNNKNDKVTICDEIKNTEHTDHEKORSVSFE 842
QY 128 GEKT-----FLYH-----SRAIEVEKTKTGSSAGLVSVATSLSHFIIYNNVSTKNDI 176
DB QEKSELISNEEYIERETTLARLTETI--FGDKNL-----FNPKMTSENKEK 888
QY 177 LHNVAQIAHCAQAKKIGSGFVATAIYGLIYYRRFAPALINDVFOVLES----- 226
DB 889 CYN-----TINELKTKYKL-PIILGNILIPYTKDNIKIKIKSI 927
QY 227 -----PEKFPPT--ELKKLIESNWEKHERCTLPYGIKLMGVKGSSETPKL 271
DB 928 EKNNKKKKKVDNPOKYFNLFFKIKNLKRNDKLYIH-----SNINKK 971
QY 272 VSRVLQMKKEPSESSVYVDOLNSANIQFMKELREKTKDSDEYTIKELDHSVEPLTV 331
DB 972 IKRAMMKKONKKCKONYAFEEDELNLNINVKQ--EKDEYDRSIRT--ENNSSAKIKD 1026
QY 332 AIKIRIRGLALQKSEVPIEPDVOTOLPRCGEIPGVGVVAGAGYDAIAVLVLENG 391
DB 1027 MNNSKNNNSQULT-NTIDIKVNNKDKAIYD-----DGHNNIQWVEIINN- 1068
QY 392 VGNFKOTLENDYFHN-----VYVVDLEQTEGVL 422
DB 1069 -----KNKENNYTHNNVQGEVLHNTKIDDDKNGEKL 1101

RESULT 39
058354 PRELIMINARY; PRT; 664 AA.
ID 058354
AC 058354
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein PH0620.
DS
OS Pyrococcus horikoshii.
OC Archaea, Euryarchaeota, Thermococci, Thermococcales, Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
EX MEDLINE=98344137; PubMed=9679194.
FA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
FA Yamamoto S., Sekine M., Baba S.-I., Kobayagi H., Hosooyama A., Nagai Y.,
FA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamaya M., Ohfuku Y.,
FA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
FA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Mauchi Y., Shituya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76 (1998).
DR EMBL; AP000003; BAA29709.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 664 AA; 77032 MW; 40AA176054479E0 CRC64;

Query Match 4.9%; Score 110; DB 17; Length 664;
Best Local Similarity 20.6%; Pred. No. 18;
Matches 73; Conservative 61; Mismatches 131; Indels 90; Gaps 18;

34 RMAAVITPKGTSLKES-----RITSSDPQANGMEWHISNTEKPREVOSRINPFL 85

Db 156 RVHSLIOAKVREIEBALRADIPFDLEIERDYGVRGFRVYAGRE--ELAGLVKPMR 212
QY 86 EATIFIVLAYIOPTFAFDE-----IIYSDPGYHSQEDTEFKTSSNGEK 130
Db 213 GGDVEI---RIQPVKRYLFEVPLKAETAIERKSIIVGLDPGI--TVGIAIDLDGNIIIS 268
QY 131 TELYHSRAITEVEK-----TGIGSSAGLVSVVATSLSH-FIPN---VISTNK 174
Db 269 VYSEKNMALSEIVRISLGHPIIVATDVNPAAGLVKARSPKQOLFPPRSLKVEEKN 328
QY 175 DILHNVAQIAHCYAOKKTIGSGFDVATAIYGLIVRRFQPALINDVFOVLESDEPEKPTL 234
Db 329 ELKKMIGISVSDHQR-----DALAAAYK--AYLRKPKL-----EHIEARL 368
QY 235 KKLISNNEKHERCTLPYGIKILMGDVYKGSFETPKLVSRVLOMKKEKPESSVYYDOLN 294
Db 369 KEL--GLMKGRNE-----IKAL---ILSGYSLGEAIKVK--TQEKPR--EVTTEREE 413
QY 295 SANLQ-FMKELREMEKRYDSDPEYIKELDHSVEPLTVAIKNIRKQLQALTQKSE 348
Db 414 SVDIKPYLKRIEEL-----ERTVEBELERENSELKAIIEQRKIIEKLERKLE 460

RESULT 40

Q9UYS7 PRELIMINARY; PRT; 643 AA.
ID Q9UYS7
AC Q9UYS7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Hypothetical protein PAB1429.
GN PAB1429.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248287; CAB50335.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 643 AA; 74576 MW; 13B4B92B5C989BEC CRC64;

Query Match 4.9%; Score 109.5; DB 17; Length 643;
Best Local Similarity 19.9%; Pred. No. 18;
Matches 73; Conservative 64; Mismatches 140; Indels 89; Gaps 17;

QY 34 RMHAIITPKGTSLKES-----RIKISSPQFANGEMWYHISNTKEKREYQSRINPFL 85
Db 136 RVHSLIOAKVREIEBALRADIPFDLEIERDYGVRGFRVYAGRE--ELAGLVKPMR 192
QY 86 EATIFIVLAYIOPTFAFDE-----IIYSDPGYHSQEDTEFKTSSNGEK 130
Db 193 GGDVEI---RIQPVKRYLFEVPLKAETAIERKSIIVGLDPGI--TVGIAIDLDGNVIA 248
QY 131 TELYHSRAITEVEK-----TGIGSSAGLVSVVATSLSH-FIPN---VISTNK 174
Db 249 VYSEKNMALSEIVRISLGHPIIVATDVNPAAGLVKARSPKQOLFPPRSLKVEEKN 308
QY 175 DILHNVAQIAHCYAOKKTIGSGFDVATAIYGLIVRRFQPALINDVFOVLESDEPEKPTL 234
Db 309 ELKKMIGISVSDHQR-----DALAAAYK--AYLRKPKL-----EHIEARL 368
QY 235 KKLISNNEKHERCTLPYGIKILMGDVYKGSFETPKLVSRVLOMKKEKPESSVYYDOLN 294
Db 349 KEY--GLMKGRNE-----IKAL---ILSGYSLGEAIKVK--AKKPREKVTERRBE 394
QY 295 SANLQ-FMKELREMEKRYDSDPEYIKELDHSVEPLTVAIKNIRKQLQALTQKSEVPIBD 354

Db 395 VDVEQYIRRIEEL-----EKTVKELERENASLRAIIIEQRKIIEKLERKLE-BFDEN 445
QY 355 VQTQL 360
Db 446 VRLKVL 451

Search completed: August 8, 2003, 09:39:44
Job time : 108 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2003, 09:28:06 ; Search time 23 Seconds

(without alignments)
883.284 Million cell updates/sec

Title: US-10-069-062-7

Perfect score: 2230

Sequence: 1 MSKAFSAPGKFLAGYLVL.....DLREQTEGVLEKEPEDYIGL 432

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

127863

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	809.5	36.3	451	1 ERG8_YEAST	P24521 saccharomyc
2	126	5.7	404	1 KIME_SCHRO	Q072780 schizosacch
3	124.5	5.6	443	1 KIME_YEAST	P072777 saccharomyc
4	123	5.5	378	1 KIME_ARATH	P46086 arabidopsi
5	119.5	5.4	526	1 VPS_BTV13	P12436 bluetongue
6	112.5	5.0	526	1 VPS_BTV13	P25179 bluetongue
7	111.5	5.0	324	1 KIME_AERPE	Q99466 aeropyrum p
8	111.5	5.0	600	1 ABR4_PLAFG	P23745 plasmodium
9	111	5.0	4092	1 DYHC_YEAST	P33022 saccharomyc
10	110.5	4.9	721	1 PRTP_HSVJ	P52385 human hepe
11	108.5	4.9	2663	1 CENE_HUMAN	Q02224 homo sapien
12	108	4.8	260	1 YD69_AQUAE	O67381 aquifex aeo
13	108	4.8	792	1 SYTB_CHUPN	O92720 chlamydia p
14	107.5	4.8	743	1 ABR4_PLAFG	P22620 plasmodium
15	107	4.8	1014	1 NEBL_HUMAN	O76041 homo sapien
16	107	4.8	1790	1 USOI_YEAST	P25386 saccharomyc
17	106.5	4.8	438	1 HISX_SHEON	O86fbl shewanelia
18	106.5	4.8	2230	1 GOG4_HUMAN	Q13439 homo sapien
19	105.5	4.7	864	1 DYH1_HUMAN	Q05193 homo sapien
20	105.5	4.7	1804	1 YFAY_YEAST	P45883 saccharomyc
21	105	4.7	1066	1 HYSA_STRPN	Q54873 streptococc
22	104.5	4.7	526	1 VPS1_YEAST	P30209 bluetongue
23	104.5	4.7	851	1 DYN1_RAT	P21575 rattus norv
24	104.5	4.7	861	1 DYN1_MOUSE	P39053 mus musculu
25	104.5	4.7	1073	1 RBSA_PLAF	P18830 plasmodium
26	104.5	4.7	4451	1 GRSB_BACR	P16888 b gramlicidi
27	103.5	4.6	436	1 PAFA_CAVPO	P70683 c platelet-
28	103.5	4.6	858	1 MUTS_BACCU	P49849 bacillus su
29	103	4.6	670	1 SSM4_SCHPO	O46677 schizosacch
30	102.5	4.6	463	1 SYC_WIGBR	O84269 wiggleswort
31	102.5	4.6	4568	1 DYHC_CARBL	Q19020 caenorhabdi
32	101.5	4.6	378	1 YN91_ANASP	P40080 anabaena sp
33	100.5	4.5	389	1 GAL1_FUSNN	O8xhdo fusobacteri

34	100.5	4.5	461	1 FTSY_AQUAE	O67066 aquifex aeo
35	100.5	4.5	782	1 YQ91_CAEEL	Q10030 caenorhabdi
36	100.5	4.5	3685	1 DMD_HUMAN	P11532 homo sapien
37	100	4.5	612	1 GIDA_MYCSE	P47619 mycoplasma
38	100	4.5	1174	1 YJ11_YEAST	P47108 saccharomyc
39	99	4.4	321	1 ABR4_PLAF	P23746 plasmodium
40	99	4.4	394	1 MK9_HUMAN	P80192 homo sapien
41	99	4.4	448	1 TRME_AQUAE	O67030 aquifex aeo
42	99	4.4	850	1 D7_DICDI	P54682 dictyosteli
43	99	4.4	891	1 SYV_PYRHO	O58052 pyrococcus
44	98.5	4.4	299	1 ARS2_AQUAE	O66674 aquifex aeo
45	98.5	4.4	497	1 AMPA_RHIME	Q92qy7 rhizodium m

ALIGNMENTS

RESULT 1
ERG8_YEAST STANDARD; PRT; 451 AA.
AC P24521;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Phosphomevalonate kinase (EC 2.7.4.2).
GN ERG8 OR YMR220W OR YM9959.02.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RN RP SEQUENCE OF 1-424 FROM N.A.
RX MEDLINE=91117228; PubMed=1846667;
RA Tsay Y.H., Robinson G.W.;
RT "Cloning and characterization of ERG8, an essential gene of
Saccharomyces cerevisiae that encodes phosphomevalonate kinase.";
RL Mol. Cell. Biol. 11:620-631(1991).
[2]
RN RP SEQUENCE FROM N.A.
RX STRAIN=5288c / AB972;
RA PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrall B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
XIII.";
RL Nature 387:90-93(1997).
[3]
RN RP -1- FUNCTION: ERG8 IS AN ESSENTIAL GENE.
CC -1- CATALYTIC ACTIVITY: ATP + (R)-5-phosphomevalonate = ADP + (R)-5-
CC -1- dihydrophomevalonate.
CC -1- PATHWAY: SECOND STEP IN ISOPRENTENYL DIPHOSPHATE FORMATION.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GMP KINASE FAMILY. MEVALONATE KINASE
CC SUBFAMILY.
[4]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
[5]
CC EMBL; M63648; AAA34596.1; -;
CC EMBL; Z49939; CA990191.1; -;
CC PIR; S57588; S57588.
CC SGD; S0004833; ERG8.
CC InterPro: IPR005203; GMPkinase_ATP.
CC InterPro: IPR005916; Pnev_kin_erg8.
CC Pfam; PF00288; GMP_kinases; 1.
CC TIGRFAMs; TIGR01219; Pnev_kin_ERG8; 1.
CC PROSITE; PS00627; GMP_KINASES_ATP; 1.

QY 156 VVATSLSHFIPNVISTNKDILHNVAQIAHCYAQKKIG-----SGFDVATAIY-GLI 206
 DB 147 VVATSLSHLAF-GNIEPPSSNSLNNKALALIBAWSTFGECIHGTFPSGIDNVAITGGLI 205
 QY 207 VYR---POPAL-----INDVQVLESDEKPEPTLKKLISGNWEKHERCTLPYGIKLL 258
 DB 206 ARKKATTAHQSAKKEFLKPKDTLSVMTD-TKPKSTKTLKVQGFELK-BR--LPVTIDSI 261
 QY 259 MGDVKGSGSETPKLVSRVLQWKKEKPESS 287
 DB 262 IDAIDGISK-----SAYLALITSESDKNSS 285

RESULT 3
 KIME YEAST STANDARD; PRT; 443 AA.
 ID KIME YEAST P07277;
 AC 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Mevalonate kinase (EC 2.7.1.36) (MK) (MWK).
 GN RARI OR ERG12 OR YMR208W OR YMR261.02.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86121728; PubMed=3323847;
 RA Kearsey S.B.; Edwards U.;
 RT "Mutations that increase the mitotic stability of minichromosomes in yeast: characterization of RARI";
 RL Mol. Gen. Genet. 210:509-517 (1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91243209; PubMed=1645230;
 RA Oulmouden A.; Karst F.;
 RT "Nucleotide sequence of the ERG12 gene of Saccharomyces cerevisiae encoding mevalonate kinase";
 RL Curr. Genet. 19:9-14 (1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=S288c / AB972;
 RA PubMed=9169872;
 RA Bommar S.; Churche C.M.; Badcock K.; Brown D.; Chillingworth T.;
 RA Connor R.; Deaman K.; Devlin K.; Gentles S.; Hamlin N.; Hunt S.;
 RA Jagals K.; Lye G.; Moule S.; Odeh C.; Pearson D.; Rajandream M.A.;
 RA Rice P.; Skelton J.; Walsh S.; Whitehead S.; Bartell B.G.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XIII";
 RL Nature 387:90-93 (1997).
 RN [4]
 RP SEQUENCE OF 1-246 FROM N.A.
 RC STRAIN=SPL;
 RA Saito A.; Kazuta Y.; Kondo H.; Tanabe T.;
 RL Submitted (OCT-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: MAY CONTRIBUTE TO THE REGULATION OF THE ISOPRENOID AND STEROL PATHWAY IN LIVING CELLS.
 CC -1- FUNCTION: RAR (REGULATION OF AUTONOMOUS REPLICATION) IS A PROTEIN WHOSE ACTIVITY INCREASES THE MITOTIC STABILITY OF PLASMIDS.
 CC -1- CATALYTIC ACTIVITY: ATP + (R)-mevalonate = ADP + (R)-5-phosphomevalonate.
 CC -1- ENZYME REGULATION: FARNESYL PYROPHOSPHATE & GERANYL PYROPHOSPHATE INHIBIT MEVALONATE KINASE BY BINDING COMPETITIVELY AT THE ATP-BINDING SITE.
 CC -1- SUBUNIT: Homodimer.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GMP KINASE FAMILY. MEVALONATE KINASE SUBFAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; X06114; CAA29487.1; -
 DR EMBL; X55875; CAA39359.1; -
 DR EMBL; 249809; CAA89923.1; -
 DR EMBL; D78165; BAA24409.1; -
 DR PIR; S05875; BABYR1
 DR SGD; S0004821; ERG12.
 DR InterPro; IPR006204; GMP kinase.
 DR InterPro; IPR006203; GMPKase ATP.
 DR InterPro; IPR006205; Mey gal kin.
 DR InterPro; IPR006206; Mey gal kinase.
 DR Pfam; PF00288; GMP kinases; 1.
 DR PRINTS; PR00959; MEVALKINASE.
 DR TIGRFAMs; TIGR00549; mevalon kin; 1.
 DR PROSITE; PS00627; GMP KINASES ATP; 1.
 KW Transferase; kinase; Sterol biosynthesis; ATP-binding.
 FT NP BIND 141 151 ATP (POTENTIAL).
 SQ SEQUENCE 443 AA; 48459 MW; 8B24052A72C97280 CRC64;

Query Match 5.6%; Score 124.5; DB 1; Length 443;
 Best Local Similarity 22.2%; Pred. No. 0.24;
 Matches 105; Conservative 72; Mismatches 154; Indels 143; Gaps 27;

QY 1 MSKAF-SAPKAPLAGIYLV--EPYDAYVTALSSMAHVIPTKGLSKSRKIKSSP 56
 DB 1 MSLPFLTAPGKVIIFGHSAYVYKPAVASAL--RTYLLIS--SSSAPDTIELLFP 55
 QY 57 QPA-NGEWEYHSS-----NTEKREVQSRINPFLATIFIVLAIQPEAPDLIELI 108
 DB 56 DISFNHKSINDPNAITEDVNSQTLAKAQ-----QATDGLSGLVS 97
 QY 109 YSDP-----GYHQEDTETKTSNGEKTFLY-----HSRAITEVEKT-----G 146
 DB 98 LLDPLLAQLSSPFYHA-----ARCFLYMFVCLCPHAKNIKFSKLTLPICAG 145
 QY 147 LGSSNGLVSVATSLSHFIPNVISTN--KILHNVAQIAHCYA--OKKI--GSGFDV 198
 DB 146 LGSSKASI--SVSLALMAVYLGILGSNDLEKISENDKRIIVQWAFIGKCHGTPSGIDN 203
 QY 199 ATATYGG--LIYRRFPQALIN-DVFOVLESDEKPEPTLKKLISGNWEKHERCTLPYGI 255
 DB 204 AVATYGNALLPEKOSHNTITNNPKFLDDP-----AI 237
 QY 256 KLMGDVKGSGSETPKLVSRVLQWKKEK-PESSVYDQNSANTQ--FMKELREMR--- 308
 DB 238 FMILTYTRIPRSTKOLVARVAVLVTEKPEVWKPIIDMAGSCALQGLEIMTKLSCKGTD 297
 QY 309 -EKYDSDEPY--IKELDHSVEPLTVAIKIRKGLQALTKQSEVPIDVDVOTQLDRQCE 365
 DB 298 DEAVETNNLELYOLELIRINHGLVLSIGVSHPELLEKLNLSD-----DLR----- 343
 QY 366 IPGCGVGV-VPGAGYDAIAVL-----VLQNVGNPKQKTLLENPPYFHVYVVDL 414
 DB 344 -----IGSYKLTGAGGGGCSLTLLRDRITQEQIDSPFKKLDQ--DPSYETFTDL 391

RESULT 4
 KIME ARATH STANDARD; PRT; 378 AA.
 ID KIME ARATH P46086;
 AC 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Mevalonate kinase (EC 2.7.1.36) (MK).
 GN ATSG27450 OR F21A20.160.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopses.

Query Match	5.4%;	Score 119.5;	DB 1;	Length 526;
Best Local Similarity	27.4%;	Pred. No. 0.68;		
Matches	75;	Conservative	34;	Mismatches 114;
			Indels	51;
			Gaps	14

RESULT 6	
VPS_BT13	
ID_VPS_BT13	STANDARD; PRT; 526 AA

CC -1- FUNCTION: THE VP5 PROTEIN IS ONE OF THE TWO PROTEINS (WITH VP2)
 J. Gen. Virol. 72:449-451 (1991).
 RT VP2.¹
 RT bluecough virus, a contrasting feature to the outer capsid protein
 RT "Sequence conservation of the outer capsid protein, VP5, of
 RA Oldfield S., Hirasawa T., Roy P.;
 RX MEDLINE=91132147; PubMed=1847179;
 RP SEQUENCE FROM N.A.
 RN NCBI_TaxId=33717;
 OS Bluecough virus (serotype 13 / isolate USA).
 OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
 OX [1]
 GN 56 OR M5.
 DE Outer capsid protein VP5.
 DT 15-DEC-1998 (Rel. 37, last annotation update)
 DT 01-MAY-1992 (Rel. 22, last sequence update)
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Created)

Query Match	5.0%;	Score 112.5;	DB 1;	Length 526;
Best Local Similarity	21.6%;	Pred. No. 2.1;		
Matches 63;	Conservative 51;	Mismatches 91;	Indels 87;	Gaps 14;

RESULT 7	
KIME_AERPE	
ID_KIME_AERPE	STANDARD; PRT; 324 AA

DT DT16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mevalonate kinase (EC 2.7.1.36) (MK).
GN MKR OR APE2439,
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcaceae;
OC Desulfurococcaceae; Aeropyrum.
OX NCBI_Taxid=56636;
OX [1]
RN SBOURCE FROM N.A.
RP
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,
RA Jin-uo K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki Y., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
RL
CC -1- CATALYTIC ACTIVITY: ATP + (R)-mevalonate = ADP + (R)-5-
CC phosphomevalonate.
CC -1- FACTOR: MAGNESTUM (BY SIMILARITY).
CC -1- SUBUNIT: Homodimer (By similarity).
CC

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE GMP KINASE FAMILY. MEVALONATE KINASE
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP000064; BAA81454.1; -
CC PIR; F72474; F72474.
CC HAMAD; MF 00217; -; 1.
CC InterPro; IPR001174; Galkinase.
CC InterPro; IPR006204; GMP Kinase.
CC InterPro; IPR006203; GMP Kinase ATP.
CC InterPro; IPR006206; Nev galkinase.
CC Pfam; PF00288; GMP kinases; 1.
CC PRINTS; PR00960; LMBPPTOTIN.
CC PRINTS; PR00959; MEVGLKINASE.
CC PROSITE; PS00627; GMP_KINASES_ATP; FALSE NEG.
CC Transferase; Kinase; ATP-binding; Magnesium; Complete proteome.
CC NP BIND 103 113 ATP (POTENTIAL).
CC
CC SEQUENCE 324 AA; 33646 MW; 8D4FC2B63C11B846 CRC64;
SQ
Query Match 5.0%; Score 111.5; DB 1; Length 324;
Best Local Similarity 22.4%; Pred. No. 1.2;
Matches 90; Conservative 58; Mismatches 146; Indels 107; Gaps 19;
QY 6 SARGKALAGVYLEPIYDAYVTALSRMAVITP--KGTSLKESRI-KISSQFANGS 62
DB 8 SARGKATIVGHEVHVAGSL-AIYALIGRLKRVTVRSKGKIVLESSLGHSHPLEGG- 65
QY 63 WEYHSSNTEKPREVOSRINPFLPATIFIVLAYIOPTAFDLIIISDPGYNQSDTER 122
DB 66 -----AAKVSFVLEPYI-----AVLRSLAAGVSVVH----- 94
QY 123 KTSNSNEKPTLYHSRAITEBEKGLGSSAGLVSVATSLSHITPNVISTNDIILNVAQ 182
DB 95 -----TLVSSGI--PFRAGLGSSA--ASMWAVA-LSYSAMGDPISADLYSVAM 140
QY 183 IAHCAVOKKIGSGFDVATAIY-GLIYRRFPALINDVQVLESDEKEPTELKLIEN 241
DB 141 EGGKIHGK-PGSDVDTIARGVGLAYRGEN-----PVDIRPGLT-- 180
QY 242 WEKHERCTLPGYIKILMGDVKGSSTPKLVSRVLQWKKEKPESSVYVDOLNSANLQPM 301
DB 181 -----GVTLLVADTGERRTDVVEHVLGIADLGEASTYIY--RAADL-IA 224
QY 302 KEIAREBEKTDSPETTYIKELHDSVEPLTVAIKIKIGLOALQKSEVPPEPVQQLD 361
DB 225 REALHAIEKGDA-----ERLGL-INNAAGLISLSIGASLEI-----TLVY 265
QY 362 RCOEIPGCVGVGAGGYDAIYLVLENQVGNFKOTLEN 402
DB 266 RMRSS-AGALGAKLTGAAGGCGVIGLFRKEGVEKGLSESVES 305

RESULT 8
ABRA PLAFG STANDARD; PRT; 600 AA.
AC P23745;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE 101 kDa malaria antigen (P101) (Acidic basic repeat antigen)
DE (Fragment).
GN ABRA.
OS Plasmodium falciparum (isolate FCR-3 / Gambia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5838;

RN [1]
RX SEQUENCE FROM N.A.
RA MEDLINE=86298794; PubMed=3042768;
RT Weber J.L., Lyon J.A., Wolff R.H., Hall T., Lowell G.H., Chulay J.D.;
RT "Primary structure of a Plasmodium falciparum malaria antigen located
RT at the merozoite surface and within the parasitophorous vacuole.";
RL J. Biol. Chem. 263:11421-11425(1988).
CC -1- SUBCELLULAR LOCATION: AT THE MEROZOITE SURFACE AND WITHIN THE
CC PARASITOPHOUS VACUOLE.
CC -1- PTM: NOT GLYCOSYLATED (PROBABLE).
CC Antigen; Malaria; Repeat.
KW
FT NON_TER 1
FT DOMAIN 133 190 8 X 6 AA REPEATS OF [VT]-N-D-[ED]-[ED]-D.
FT REPEAT 143 148
FT REPEAT 149 154
FT REPEAT 155 160
FT REPEAT 161 166
FT REPEAT 167 172
FT REPEAT 173 178
FT REPEAT 179 184
FT REPEAT 185 190
FT DOMAIN 591 >600 TANDEM REPEATS OF K-E(2,3).
FT NON_TER 600 600
SQ SEQUENCE 600 AA; 69293 MW; F374614F24AAC938 CRC64;
Query Match 5.0%; Score 111.5; DB 1; Length 600;
Best Local Similarity 21.3%; Pred. No. 3;
Matches 83; Conservative 56; Mismatches 110; Indels 141; Gaps 19;
QY 126 SNGEKTFLYHSRAITE-----VEKTGLGSSAGL-----VSVAATSLSH-FIPNVISTN 173
DB 239 SNDLINFNLNGKEIGENLNTITKNMLGDKNNLESFPLDQNMKONLINVEFTLNNTKS 298
QY 174 -----KDLIYNAQIAH-CYAOKKIGSGFDVATAIYGLIYRRFPALINDVQVLESBP 227
DB 299 VLNKIKDILRLLYKAYVSYKRRKAQE-----KGLP-----EPVTN----- 335
QY 228 EKEPTELKLIENWEKHERCTLPGYIKILMGDVKGSSTPKLVSRVLQWKKE----- 281
DB 336 EBYVELLKKGI-----LDMGKILFSKTK--SLKGLKKNKLPKKKENONAVD 381
QY 282 -----KPESSVYVDOLNSANLQ--FMKELRE----- 306
DB 382 TKSMEBPVKAQAPALRGVPEPDSINMNSINNVMDIOPFEKELIENNTNPVVPPTQSK 441
QY 307 -----WEKTDSPETTYIK-----LDHVEPLTVAIK-----NIRKG 339
DB 442 KKNKNETVSGMDENFDNHEVFKEEYIYDENDDMEVKKGIGVTLKKEPPLKNGVSET 501
QY 340 LQAL-----TQSEVPFIEPDVQQLDRCOEIPGCVGVGAGGYDAIYLVLENQVGN 394
DB 502 IYLIHGNKDKKHITAIINNDIOLIK-----QELQAIYVELMAYTNNGNKYIQO----- 549
QY 395 FKQKTLNPDYFHNAYVWDLBEQTEGVLEE 424
DB 550 PQQNTLEN-DVLANQETBEEMEQVEATIKQ 578

RESULT 9
DYHC YEAST STANDARD; PRT; 4092 AA.
AC P36022;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Dysein heavy chain, cytosolic (DYHC).
GN DYNI OR DHCI OR YKR054C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
SEQUENCE FROM N.A.


```

CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES PRTP FAMILY.
CC
CC This SMS-S-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U43400; AAC54702.1; -.
CC PIR; T41942; T41942.
CC InterPro: IPR000501; Proc_transpt.
CC Pfam: PF01366; PRTP; 1.
CC Capsid assembly.
CC SQUENCE 721 AA; 83166 MW; BE5BED64E46EC9CC CRC64;
CC
Query Match
Best local Similarity 20.5%; Score 110.5; DB 1; Length 721;
Matches 100; Conservative 74; Mismatches 197; Indels 117; Gaps 23;
CC
QY 13 LAGGYVLEP-IDAYVYALSSMAHAVTP-KGTSLKESRIKISSPOFANGEMEYHISN 70
DB 171 LCGVCVMKREVESEFTTCLCKYEELSLVNOCKSIR---KILACKFCNHLTETMVSN 226
QY 71 TEKREVSQINPLEATFIVLAYI-OPEAFDLEIIYSDPGVHSEDETKTSSN-- 127
DB 227 LKXNDVIEKDDPFRKQGLVKEVMAKTNITQOC-LYSKRP-HLQEMENLINFDF 283
QY 128 -----GKPTLYHSRAITE--VEKTLGSS-----AGLVSVVATSLSHFIPN 168
DB 284 SKIPDTYLSSEFTYMSK-ISESVIQASITLQMLCHSLVADLQNEISKFLYGTID 342
QY 169 VISTKDIHLNVAQIAHCAQKIGSGFDVATLYGLIYRRQPLINDVFOVLESDEP 227
DB 343 VNFNEE--NVTNDDKLY---IGSRF-----ISPCRLVDITIVNSIGNLEEDV 386
QY 228 -----EKPPTEKXLISSNMEKHERCTLPYGIKLMGD-----VKGSSETPKX- 271
DB 387 FTKLAEBDELQYIKITLNLLENLSAHTVPKXVTHSMODHNLQOEHIIRKAYYQKIS 446
QY 272 ---VSRVLQWKKEPRESSV-----VYDOLNSANLOFMKELEMERKYSDE 316
DB 447 ESGYSVLMCIKX--GEALINKLMNTILGNHIFESLSKMMNFAHQQLSLENFSADFP 504
QY 317 TY-----IKELDSHVEP-----LTVAIKIRKGLQALTQKSEVPIDPVO 356
DB 505 TYDDHLIYIKNNILSKLPELLEPNLSQEWYRLLTGPLSNYHTASFPSSNISMAVACDVA 564
QY 357 TOLLDRCCOIRPGCVGVPPAGSYDAIVAVLVLENQVGNFKQKTLNPDYFHNHYVWDL 416
DB 565 DFLPHKEDJAKCVBGTI-----YPENMLCTYKKEFFN-----DGLHNIN--DMQR 609
QY 417 QTEGVLE 424
DB 610 QMNMFIRE 617

```

```

RESULT 11
CENE HUMAN
ID CENE HUMAN STANDARD; PRT; 2663 AA.
AC 002224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_Taxid=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539 (1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RL EMBO J. 14:918-926 (1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=96437347; PubMed=9763420;
RA Chan G.K.T., Schar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and HUBB1.";
RL J. Cell Biol. 143:49-63 (1998).
CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC
CC This SMS-S-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; Z15005; CAA78727.1; -.
CC PIR; S28261; S28261.
CC HSP; P17119; 3KAR.
CC Genew; HGNC:1856; CENPE.
CC
CC GK; Q02224; -.
CC MIM; 117143; -.
CC GO; GO:0005699; C:kinetochore; TAS.
CC GO; GO:0005634; C:cnucleus; TAS.
CC GO; GO:0008350; F:kinetochore motor activity; TAS.
CC GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
CC GO; GO:0007079; P:mitotic chromosome movement; TAS.
CC GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
CC InterPro: IPR001752; kinesin_motor.
CC Pfam: PF00225; kinesin; 1.
CC PRINTS; PR00380; KINESINHEAVY.
CC SMART; SM00129; KISC; 1.
CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
CC PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
CC Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
CC Cell cycle; Centromere.
CC KINESIN-MOTOR.
CC COILED COIL (POTENTIAL).
CC GLOBULAR (POTENTIAL).
CC NP BIND 86 93 ATP (BY SIMILARITY).
CC SQUENCE 2663 AA; CEFCL3880C8C8B8 CRC64;
CC
Query Match
Best local Similarity 4.9%; Score 108.5; DB 1; Length 2663;
Matches 84; Conservative 67; Mismatches 161; Indels 93; Gaps 16;
QY 47 KESRIKISSPOFANGEMEYHISNTEKREVSQINPLEATFIVLAYIOPTAFDDEI 106
DB 852 KKAQFDSLSGLAKTSLSTYQEOLEKTRVQERLNE-MEQLKEQLERNRDSPLQVEREK 910
QY 107 IYSDPGVHSEDETKTSSNGEKTFLYHSRAITEVEKTLGSSA----- 151

```

```

Db      911 TLTEKLOQTLEEVTLTQ---EKDLDKLOQESLQIERDQLKSDIHDVTNNMIDTQEOQLR 967
Qy      152 -GLVSV---VATSLSHPIPNVISTNKQILINVAQIACHVAKKQIG-----SGFD 197
Db      968 NALSELKHQHOETINTLKKSISEVRNLMHEENTSETODEFOQKAVGIDKDKDLEAKTQ 1027
Qy      198 VATA-----IYGLIVRRFQPALINDVFOYLES---DPEKFPTELKLIIESN 241
Db      1028 TLTAADVKNONEIEQQRKIFSLIQEK-----NELQOMLESYIAKEQDLDKTDKENIEMT 1080
Qy      242 WEEKERCTLPYGTILMGDVKGGSETPVLVSRVLOMK---EKPESSVYVDOLSNANL 298
Db      1081 IENDEE-----RLTL-----GDELKKQOEIVAQEKHAIKKEGELSRICDRLAEVEE 1127
Qy      299 QPMKELREMRREKYS--DPEYTIKELHDSVEPLTVAIKNRKQALQOKSEVPLEPVO 356
Db      1128 KTKESQQLQEQKQOOLNVAQESMSQKKIN---ETELKNEL---KNKELTLE-HWE 1178
Qy      357 TQLDRCOEIPGCVGVVPGAGYDAIVLLENQGNFKQKYLE 401
Db      1179 TERLELAQKL-----NENVEVKSITKERVKLDELQKSPF 1213

```

RESULT 12
YD69_AQUAE STANDARD; PRT; 260 AA.

```

ID      16-OCT-2001 (Rel. 40, Created)
AC      067381;
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical protein AQ_1369.
GN      AQ_1369.
OS      Aquifex aeolicus.
OC      Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
CX      NCBI_TaxID=63363;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=VF5;
RX      MEDLINE=98196666; PubMed=9537320;
RA      Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA      Graham D.E., Overbeek R., Sneed W.A., Keller M., Aubay R.,
RA      Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT      "The complete genome of the hyperthermophilic bacterium Aquifex
RT      aeolicus";
RL      Nature 392:353-358(1998).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AE000737; AAC07356.1; -
DR      PIR; B70419;
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 260 AA; 30206 MW; B5D3F66F37C69B3 CRC64;

```

```

Query Match      4.8%; Score 108; DB 1; Length 260;
Best Local Similarity 24.8%; Pred. No. 1.6;
Matches 58; Conservative 26; Mismatches 68; Indels 82; Gaps 13;
Qy      205 LIVR-----RQPLINDVFOYLESDEPKFTELKLIIESWEEKHERCTLPYGTILMG 260
Db      85 LVIRIRPEKEPEKEL--GIFQSV---EAMGAFSLTALEHQMEE---VPKNVYIYHA 133
Qy      261 D-VKGSSTPLVSRVLOMKKEKPESSVYVDOLSNANLQPMKELREMRK----- 310
Db      134 DFVEGGA---KLIAAI-----KTEBEGISTYDQL-----KLEEMKKKAVPRVAV 175
Qy      311 YDSDPEYIKELHDSVEPLTVAIKNRKQALQOKSEVPLEPVOQLDRCOEIPGCV 370

```

```

Db      176 YSSDVLTYIKD-----IYPPDVQSKAVYIARBIARET 206
Qy      371 GGV--VPGAGYDAIVLLENQGNFKQKYLEND-----DYHNYVWYDLE 415
Db      207 GSAPLEELIKGIYGVDTSTLEKGL-ELIEKLLQNVKLPDGKENVLKPYWPLE 259

```

RESULT 13

SYFB_CHLPN STANDARD; PRT; 792 AA.

```

ID      09Z7W0; Q9JSD3; Q9K2D4;
AC      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20)
DE      (Phenylalanine--tRNA ligase beta chain) (PheRS).
GN      PHET OR CPN0594 OR CP0154.
OS      Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC      Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
CX      NCBI_TaxID=83556;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CWL029;
RX      MEDLINE=99206606; PubMed=10192388;
RA      Kaiman S., Mitchell W., Marathe R., Lammell C., Fan J., Hyman R.W.,
RA      Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT      "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";
RL      Nat. Genet. 21:385-389(1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=AR39;
RX      MEDLINE=20150255; PubMed=10684935;
RA      Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA      White O., Hickey E.K., Peterson J., Utterback T., Berry K., Baes S.,
RA      Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA      Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA      Eisen J., Fraser C.M.;
RT      "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
RT      pneumoniae AR39";
RL      Nucleic Acids Res. 28:1397-1406(2000).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=J138;
RX      MEDLINE=20330349; PubMed=10871362;
RA      Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA      Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT      "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT      from Japan and CWL029 from USA";
RL      Nucleic Acids Res. 28:2311-2314(2000).
CC      -----
CC      -1- CAPALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC      diphosphate + L-phenylalanyl-tRNA(Phe).
CC      -1- COPACTOR: Binds 2 magnesium ions per tetramer (By similarity).
CC      -1- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC      similarity).
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC      -1- SIMILARITY: Belongs to the phenylalanyl-tRNA synthetase beta chain
CC      family. Subfamily 1. tRNA-binding domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AE001644; AAD18733.1; -
DR      EMBL; AP002176; AAF38035.1; -
DR      EMBL; AP002547; BAA98801.1; -
DR      PIR; D81608;
DR      PIR; G86564; G86564.

```

DR PIR; H72058; H72058.
 DR TIGR; C0154; -.
 DR HAMAP; MF 00283; -; 1.
 DR InterPro; IPR005146; B3_4.
 DR InterPro; IPR005147; B5.
 DR InterPro; IPR005121; Fdx-AnticB.
 DR InterPro; IPR004532; phet bact.
 DR InterPro; IPR002547; tRNA_bind.
 DR Pfam; PF03483; B3_4; 1.
 DR Pfam; PF03484; B5_1.
 DR Pfam; PF03147; FDX-ACB; 1.
 DR Pfam; PF01588; tRNA_bind; 1.
 DR TIGRFAMs; TIGR00472; phet bact; 2.
 DR PROSITE; PS50886; TRBD; 1.
 DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KM Metal-binding; Magnesium; RNA-binding; tRNA-binding;
 KM Complete proteome.
 FT DOMAIN 39 154 TRNA-BINDING.
 FT METAL 456 456 MAGNESIUM (BY SIMILARITY).
 FT METAL 462 462 MAGNESIUM (VIA CARBONYL OXYGEN) (BY
 SIMILARITY).
 FT METAL 465 465 MAGNESIUM (BY SIMILARITY).
 FT METAL 466 466 MAGNESIUM (BY SIMILARITY).
 FT CONFLICT 171 171 G -> A (IN REF. 3).
 FT CONFLICT 577 577 C -> Y (IN REF. 1).
 SQ SEQUENCE 792 AA; 87732 MW; 852C86BF0F07486F CRC64;

Query Match 4.8%; Score 108; DB 1; Length 792;
 Best Local Similarity 19.7%; Pred. No. 7.7;
 Matches 75; Conservative 69; Mismatches 133; Indels 104; Gaps 18;

QY 36 HAVITPKGTSLKESRIKISSPOFANGEMWEYHISNTEKPREVSRINPLEATIFVLAY 95
 DB 221 YVIT--GISAQSPFIKL-----OESLQALKQKP-----INAVDITNYIMLSL 262
 QY 96 IOPTEAFDEILIIYSDPGYHSEDTETKTSSNGEKTFLVHSRAITEVEKT--GLSSAG- 152
 DB 263 GQPLHAYDASHVALDLRVEKLTSTPESTLTILNGETVLTPSGVAVVDDHSLGLGGMGA 322
 QY 153 ----LVSVAATSL--SHFTPNVISTNKDILHVAQIAHCYAKKIGSGFDVAATAYGLI 206
 DB 323 KAPSFQETTTTIVYKAYFLPEALRASQKLLPIPSBSAVRFR---GIDQNVV--- 373
 QY 207 VYRRFOPALINDVFOYLESDEPKT-----PTELKKLIESNMEBK 246
 DB 374 ----PALQAHVILYELFEPEKTSPIYSSGEICELKEVALRPETLQRIIGKS----- 422
 QY 247 ERCTLPYGIKLMGVK--GQSETPK--LVSRLQWKKEKPESSVVD--QLNSANLQ 299
 DB 423 ----FSIEILISQKLSIGSFSTTPOETSLVVPYRHDINEIDLVEICRTESMWIE 476
 QY 300 FMEKLEMEKXKDSDEP-----YIKELDHSEVPLVAIKNRKGLQALTOKS 347
 DB 477 TONPVSCYPIYKLRKETAGFLANAGLOEFPFDL--LDPEVAL--TRK-----EXE 525
 QY 348 EVPIEPDVOTQLDRCOEIG 368
 DB 526 EISLQSKHTIVL-RSSLIPG 545

RESULT 14
 ABRRA_PLAFC STANDARD; PRT; 743 AA.
 ID ABRRA_PLAFC
 AC P22620;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE 101 kDa malaria antigen (P101) (Acidic basic repeat antigen).
 GN ABRRA.
 OS Plasmodium falciparum (Isolate Camp / Malaysia).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_Taxid:5835;
 DT (1)

RP SEQUENCE FROM N.A.
 RX MEDLINE=88298794; PubMed=3042768;
 RA Weber J.L., Lyon J.A., Wolff R.H., Hall T., Lowell G.H., Chulay J.D.;
 RT "Primary structure of a Plasmodium falciparum malaria antigen located
 at the merozoite surface and within the parasitophorous vacuole.";
 RL J. Biol. Chem. 263:11421-11425(1988).
 CC -1- SUBCELLULAR LOCATION: AT THE MEROZOITE SURFACE AND WITHIN THE
 CC PARASITOPHOUS VACUOLE.
 CC -1- PTM: NOT GLYCOSYLATED (PROBABLE).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; J03902; AAA29462.1; -.
 DR PIR; A29232; A29232.
 DR Antigen; Malaria; Repeat.
 KM DOMAIN 226 273 8 X 6 AA REPEATS OF [VT]-N-D-[ED]-[ED]-D.
 FT REPEAT 226 231
 FT REPEAT 232 237
 FT REPEAT 238 243
 FT REPEAT 244 249
 FT REPEAT 250 255
 FT REPEAT 256 261
 FT REPEAT 262 267
 FT REPEAT 268 273
 FT DOMAIN 674 731
 SQ SEQUENCE 743 AA; 86622 MW; FBF0B0FB807D922C CRC64;

Query Match 4.8%; Score 107.5; DB 1; Length 743;
 Best Local Similarity 21.0%; Pred. No. 7.6;
 Matches 82; Conservative 50; Mismatches 117; Indels 141; Gaps 17;

QY 126 SNGEKTFYHSRAITE--VEKTGLSSAGIVSVATSL-----LSHFIPN 168
 DB 322 SNDLINFANQGEKEIENLINTKNLGDKNLSPDELMKLNINYEITLIDNLKTS 381
 QY 169 VISTNKDILHVAQIAH-CYAOKKIGSGFDVAATAYGLIYVRRFOPALINDVFOYLESDP 227
 DB 382 VLNKDKDILLRLLYAYVSYKRRKAQ--KGLP-----EPVTN----- 418
 QY 228 EKFPELKKLIESNMEHGERCTLPYGIKLMGVKSGSETPKVSRLQWKE----- 281
 DB 419 EYVEELKKG1-----LDWGIKLFSRYK--SLKDKNKKLPKKGEDNQAVD 464
 QY 282 ----KPESSVVDQLNSANLQ-----FMKELRE----- 306
 DB 465 TKSMEPPKVAQPALRGVETEDSNINMSINNWADEIDFEKELIENNTNPNVPPQSK 524
 QY 307 ----MKKYSDPEYIKE-----LDHSVEPLVAIK-----NIRKG 339
 DB 525 KKNKNETVSGMDENDPNPENYFKEEYYDNDWEVVKKIGVTLKKEEPLKGNVSET 584
 QY 340 LQAL-----TQKSEVPIEPDVOTQLDRCOETPGCVGVVPGAGGYDAIAVLLENQGN 394
 DB 585 IKLHLGNKDKKHIEALINNDIQIIRK--DELQAIYELMAYTNGNKNIQOI----- 632
 QY 395 FKQKLTENDYFHHVYVWDLSEQTEGVLEE 424
 DB 633 FQONITEN-DVLNQETEEEMEKQVAYTKQ 661

RESULT 15
 NEBL_HUMAN STANDARD; PRT; 1014 AA.
 ID NEBL_HUMAN
 AC 076041; Q9UITG;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)

Db 346 NKGKPLVEFVETPSYQASKEAQKQKQSEKYYKEDFEKEIKGRSLDIDKTPR 396

RESULT 16
USO1_YEAST STANDARD; PRT; 1790 AA.

AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Intracellular protein transport protein USO1.
GN USO1 OR INT1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Saccharomycetes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae;
OC NCBI_Taxid=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, uso1, is required for intracellular
RT protein transport in Saccharomyces cerevisiae.";
RL J. Cell Biol. 113:245-260(1991).
RN (2)
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -1- SIMILARITY: BELONGS TO THE VDP/USO1/YDL047C FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X54378; CAA38253.1; -
CC EMBL; L03188; AAB00143.1; -
CC EMBL; U53668; AAB6659.1; -
CC SGD: S0002216; USO1.
DR InterPro: IPR002017; Spectrin.
DR InterPro: IPR006955; USO1_p115_C.
DR InterPro: IPR006953; USO1_p115_head.
DR Pfam: PF04871; USO1_p115_C; 1.
DR Pfam: PF04869; USO1_p115_head; 1.
KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (REPETITIVE).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 847 847 G -> E (IN REF. 2).
FT CONFLICT 924 924 E -> K (IN REF. 2).
FT CONFLICT 1253 1253 V -> I (IN REF. 2).
FT CONFLICT 1319 1319 I -> V (IN REF. 2).
FT CONFLICT 1461 1461 N -> S (IN REF. 2).

FT CONFLICT 1581 1581 G -> S (IN REF. 2).
FT CONFLICT 1600 1600 I -> V (IN REF. 2).
FT CONFLICT 1661 1661 R -> S (IN REF. 2).
FT CONFLICT 1772 1772 D -> DEEDDE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B2169FD4818 CRC64;

Query Match 4.8%; Score 107; DB 1; Length 1790;
Best Local Similarity 20.2%; Pred. No. 29;
Matches 87; Conservative 67; Mismatches 166; Indels 110; Gaps 16;

QY 13 LAGGLVLEPIPDVATVTLSSRMNAVITPKGTSLKESIKISSPPANGEMWYHSSNTE 72
DB 967 LANNYKMDQAEENESIKAVE-----SKNESSIQLSNLQ-----NKIDSMSQ 1008
QY 73 KREYQSNINPLEATITVLAIGPTAPADEIITVDPG---YHSQ-----EDPTKTT 124
DB 1009 EKKNFOIR-RGSIEKNISQLKTTISDLHQTBEIISKSDSKDEYSQISLKEKLETA 1067
QY 125 SNGEKTEFLYSHRAITEVEKTLGSSAGLVSVATSLSHPIPNVISTNKDILHNVAQIA 184
DB 1068 TANDENV-----NKISELTKREBLEAEL---AAVKNLKNLELTLETSEKALKEVENE 1119
QY 185 HCVAQKTIQSGPDVATVLAIGLVRRFPQALINDVQVLESDDPEKFTPELKLIE--SNW 242
DB 1120 EHLKEEKTQLEKEATET-----KQQLNSLRANLSLKEKHEDLAAQLKKEEQIANK 1171
QY 243 EEKHERCTLPYGIKLMGVKGSSE-----TPKLSRVLYQMKKEKPESSVYVDLN 294
DB 1172 EKQYNE-----EISQINDEITSTQGENSIRKKNDELGEVAKAKSTEBEGSNLKKSEID 1226
QY 295 SANIQFMKELRMRERKYSDDPEYIKELDSVEPLTVAKIRKGLQALYQKSEVPISPD 354
DB 1227 ALNLQ-----IKELKKQNETN-EASLLESIKSVESETVXIK----- 1261
QY 355 VQTQLDRCQETPGCVGVVPGAGGYDAIIVLLENQVGNKQKTLNPDYFHNVYWDL 414
DB 1262 ---ELQDCN-----FKKEVSELEDKXKASED--KNSKYTEL 1294
QY 415 EEQTEGLVE 424
DB 1295 QSESEKIVE 1304

RESULT 17
HISX_SHEON STANDARD; PRT; 438 AA.
AC Q8EBF1;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Histidinol dehydrogenase (EC 1.1.1.23) (HHD).
GN HISD OR SO2073.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_Taxid=70863;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Meche B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouli H., Gill J., Utecherback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
CC -1- FUNCTION: Catalyzes the sequential NAD-dependent oxidations of L-
CC histidinol to L-histidinolaldehyde and then to L-histidine (By

CC similarity).
 CC -1- CATALYTIC ACTIVITY: L-histidinol + 2 NAD(+) + H(2)O = L-histidine
 CC + 2 NADH.
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -1- PATHWAY: Histidine biosynthesis; ninth (last) step.
 CC -1- SIMILARITY: Belongs to the histidinol dehydrogenase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE015649; AAN55120.1; -
 CC TIGR: SO2073; -
 CC HAMAP: MF_01024; -; 1.
 CC Pfam: PF00815; Histidinol_dh; 1.
 CC Prodom: PD002680; Histidinol_dh; 1.
 CC TIGRFAMs: TIGR00069; hsd; 1.
 CC PROSITE: PS00611; HISOL_DEHYDROGENASE; 1
 CC Histidine biosynthesis; Oxidoreductase; NAD; Metal-binding; zinc;
 CC Complete proteome.
 CC ACT_SITE 332 BY SIMILARITY.
 CC ACT_SITE 333 BY SIMILARITY.
 CC METAL 265 ZINC (BY SIMILARITY).
 CC METAL 268 ZINC (BY SIMILARITY).
 CC METAL 366 ZINC (BY SIMILARITY).
 CC METAL 425 ZINC (BY SIMILARITY).
 CC SEQUENCE 438 AA; 46296 MW; D7D034A1C5ABE3 CRC64;
 Query Match 4.8%; Score 106.5; DB 1; Length 438;
 Best Local Similarity 22.3%; Pred. No. 4.2; Mismatches 112; Indels 121; Gaps 20;
 Matches 81; Conservative 49; Nonconservative 112; Indels 121; Gaps 20;
 QY 20 LEPIYDAYVTLSSRMHVAVITPKGTSLSKRSIKISSPQFANGEMWYHISNTEKPR----- 75
 DB 29 LEQSVRAIVDAVASHGDAI-----KEFNQKFGARLANIS-----SANSNDNRLESH 76
 QY 76 EVQ--SHINPLENTITVLAIVIOPTAFDELEIITISYDPCYHSGE-----DTEKTSNN 127
 DB 77 EIEASARVSPDLKKAIAQAMANIDV-----FHSAQCFRPIDIETQAGVR 121
 QY 128 GEKTFLYHSRAITEKTKGL---GSSAGLVSVVATSLSHLPNNTS-----T 172
 DB 122 CE-----LSEPIEKVGLYIPGSSAPLISIV---LMLAPITACEQEVLSPPPI 170
 QY 173 NKDILH--NVAQIAHCYAKKIGSGFDVATAYGL---IYRRFOPALINDVEQVLES 226
 DB 171 NDALIVYANVGGITEIY---GVGQAQALALAFGETETIPVDKIFGPG----- 215
 QY 227 PEKPEPELAKLIESNWEKHERCTIPYGIKILMGDVKGSGSETPKLVSVYLQWKKEKPEES 286
 DB 216 -NRYVTEKRVLSQD---GRCTVSI-----DMPAG-----PSRV 245
 QY 287 SVVVDQANSLQFPEKELREMEKREKSDPEYIKELSHSVPLTVAIKINIKGIALTLQK 346
 DB 246 LVIND--SDANARQPI--AADLISQAEHGSDCVIILVDTSL--PLAQAVQALSKQLALPR 300
 QY 347 SEV 349
 DB 301 QEI 303
 RESULT 18
 GOG4 HUMAN STANDARD; PRT; 2230 AA.
 ID GOG4 HUMAN 013439; 013270; 013654; 014436;
 AC 013439; 013270; 013654; 014436;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Golgi autoantigen, golgin subfamily A 4 (Trans-Golgi p230) (256 kDa

DE golgin (golgin-245) (72.1 protein).
 GN GOLGA4.
 OS Homo sapiens (Human).
 OC Eukaryota; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OK NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=96215236; PubMed=8626529;
 RA Erlich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;
 RT "Molecular characterization of trans-Golgi p230: a human peripheral
 RT membrane protein encoded by a gene on chromosome 6p12-22 contains
 RT extensive coiled-coil alpha-helical domains and a granin motif.";
 RL J. Biol. Chem. 271:8328-8337(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Seelig H.P.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 131-2230 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=96125112; PubMed=8537393;
 RA Fritzel M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;
 RT "Molecular characterization of golgin-245, a novel Golgi complex
 RT protein containing a granin signature.";
 RL J. Biol. Chem. 270:31262-31268(1995).
 RN [4]
 RP SEQUENCE OF 524-672 FROM N.A.
 RC TISSUE=Gastric fundus;
 RA Balague C.;
 RL Thesis (1994), Instituto municipal de investigación medica, Spain.
 CC -1- FUNCTION: May play a role in vesicular transport from the trans-
 CC Golgi.
 CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CYTOPLASMIC FACE OF THE
 CC GOLGI MEMBRANE.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=4;
 CC Comment-Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=Q13439-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q13439-2; Sequence=VSP_004272, VSP_004273;
 CC Name=3;
 CC IsoId=Q13439-3; Sequence=VSP_004274;
 CC Name=4;
 CC IsoId=Q13439-4; Sequence=VSP_004275;
 CC -1- DISEASE: ANTIGEN IN THE AUTOIMMUNE DISEASE SJOERGEN'S SYNDROME AND
 CC IN HEPATITIS B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U41740; AAC50434.1; -
 CC EMBL: X82834; CAAS8041.1; -
 CC EMBL: U31906; AAC51791.1; -
 CC EMBL: X76942; CAAS4261.1; -
 CC Genew: HGNC:4427; GOLGA4.
 CC MIM: 602509; -
 CC GO: GO:0005802; C:Golgi trans-face; TAS.
 CC GO: GO:0016192; P:vesicle-mediated transport; TAS.
 CC InterPro: IPR000237; GRIP_domain.
 CC Pfam: PF01465; GRIP; 1.
 CC Golgi stack; Antigen; Coiled coil; Alternative splicing.
 FT DOMAIN 133 237 COILED COIL (POTENTIAL).
 FT DOMAIN 276 1011 COILED COIL (POTENTIAL).
 FT DOMAIN 1033 1214 COILED COIL (POTENTIAL).
 FT DOMAIN 1259 2152 COILED COIL (POTENTIAL).
 FT VARSPIC 2154 2185 TPKGGMVHTVDVSLFSEPFETFLRVLFERY -> HLTKV

FT STRAND 561 566
 FT TURN 568 569
 FT STRAND 570 574
 FT STRAND 584 590
 FT TURN 591 592
 FT TURN 598 599
 FT STRAND 603 607
 FT HELIX 610 622
 FT TURN 623 624
 FT STRAND 626 627
 SQ SEQUENCE 864 AA; 97407 MW; 782D86B572FFED0F CRC64;

Query Match 4.7%; Score 105.5; DB 1; Length 864;
 Best Local Similarity 22.6%; Pred. No. 13;
 Matches 67; Conservative 46; Mismatches 116; Indels 67; Gaps 12;

QY 76 EVQSGINPELEKTIYVLAAYIOPTAFPLEIITYD-----PGY-----HSGEDPETHKS 125
 DB 188 KVAKEVDPOGQRTIGVITKLDMDGTDARDVLEKLLPLRRGYIGVNRSGKIDGKDD 247
 QY 126 SNG-----EKTLYHSRAITEVEKTL-----GSSAGLVSVATSLIS- 163
 DB 248 ITAALAAERKFFLSHPSRHLADRMGTPLYLQKVLNQULTNHRTLLPGLRNKLGSLSTI 307
 QY 164 -----HFLP-NVISTNKDILHNVAQIAHCYAQKIGSGFDVATAYGLVYRFPQA 214
 DB 308 EKEVEYNFRDPDPAKTKALLQWVQFADVFEKRISSGQIDT-----YELSGA 360
 QY 215 LINDVFOYLEDPEKFFELKLLIESNBEKHERCTLPYGIKLMGVKSGSETPKLVSR 274
 DB 361 RINRFH-----ERFPELVYM--EFDEKELREISYAIKNING-IRTGIFTPDMAFE 410
 QY 275 VLWQKKEKPEBS-----SVVYDOLNSANLQFMKELRE--MREKYDSDPEYIKK 321
 DB 411 TIVKQVKIKRPPCLKCVDMVISELITVROCTKLDQYPRLEMERIVTTHIRE 466

RESULT 20
 FEAT_YEAST STANDARD; PRT; 1804 AA.
 ID_YEAST
 AC P43563;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical 207.6 kDa protein in SWC1-SEC4 intergenic region.
 GN YFL007M.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / AB972;
 RX MEDLINE=95400292; PubMed=7670463;
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
 RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
 RA Yamazaki M., Tashiro H., Eki T.;
 RT "Analysis of the nucleotide sequence of chromosome VI from
 RT Saccharomyces cerevisiae."
 RL Nat. Genet. 10:261-268(1995).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL, D50617; BA009231.1; -.
 DR FPI; S56247; S56247.
 DR SGD; S000187; BLM3.

KM Hypothetical protein; Transmembrane.
 FT TRANSMEM 143 163 POTENTIAL.
 FT TRANSMEM 373 393 POTENTIAL.
 FT TRANSMEM 404 424 POTENTIAL.
 FT TRANSMEM 452 472 POTENTIAL.
 FT TRANSMEM 564 584 POTENTIAL.
 FT TRANSMEM 680 700 POTENTIAL.
 FT TRANSMEM 912 932 POTENTIAL.
 FT TRANSMEM 1389 1409 POTENTIAL.
 FT TRANSMEM 1627 1647 POTENTIAL.
 FT TRANSMEM 1671 1691 POTENTIAL.
 SQ SEQUENCE 1804 AA; 207614 MW; D4C9E010517EEB58 CRC64;

Query Match 4.7%; Score 105.5; DB 1; Length 1804;
 Best Local Similarity 18.5%; Pred. No. 37;
 Matches 113; Conservative 84; Mismatches 184; Indels 231; Gaps 31;

QY 17 YLVLEPIYDAVYALSSRHMAVITPKGTSLESRIK--ISSPOF-----ANGEME----- 64
 DB 651 FIVMDKLYRHVITVNLVSLVSKLDMNDPTNLISNLINGIVSIAAFIPIDGLGDDYISFB 710
 QY 65 -----YHI-----SSTKPKR-----VQSRINPLELTIPIV----- 92
 DB 711 SDTLPVQHFYHFKCGESSSTFPVDELLANNAKASTVVFQSMKLYVEKIFQVADVDL 770
 QY 93 -----LAVIOP-----EAFDLEIITY-----SDPGYHSGEDPETHK-- 124
 DB 771 EDSLVTKINOTTMLOESMDKIRYFASLIQRNFWSDSKREKDPNT-----ELVTIP 824
 QY 125 -----SSNG-----EKTFLYHSRAITEVEKTLGSSAGLVSVATSLISLRIPNV- 170
 DB 825 LAALVRNNGSKSLVRLTFH-----IKEQIRKAG-----SVSTSEIQQRDVGLVYL 875
 QY 171 STNKDILHNVAQIAHCYAQKI-----GSGFDVATAYGLVYRFPQALINDV 219
 DB 876 TALNDVLRQCHESLLEYSDELITPMKLYLDVNTNPPDLVITSI--VHSALATCTTEI 932
 QY 220 FQV-----LESDEKPEPTELKLLISNM-----EKGK 247
 DB 933 TDCRLFPEDSKIPKDKMGLOQFPRFPD--KQHSQMHVPSDSDEITLSISLESSE 989
 QY 248 RCTLPYGIKLMGVKSGSETPKLVSRVLOMKKEKPEBSVYVD-----QLNSANLQPMK 302
 DB 990 YCI--NNVEBELMKAPRHSDEYGMIOKYVLAWHTLSSSLFLPDPFMYKRTQSLSVRE 1047
 QY 303 E---LREMRKRYDSDEPYIYELDHSVPLT-----VAIKIRKQLQ-----LTOK 346
 DB 1048 KLILKIRREN-NCDPQ---ELDIDIEQIRSGKODEDYISKDIEAGLNGVSDVQJLR 1102
 QY 347 SEVPLEPDVQTLDRCOEIRPGCVGVPGAGYDATA-----VLVLENGVNF 395
 DB 1103 DEFPDELIVDESV--SEMPSGVNTPIAGTHGIDNSAMSSDLAFRLDIDTYTCNYFNT 1159
 QY 396 KQKTLNPDY-----FHNVY-----WVLEBQOT 418
 DB 1160 TEELQLQPOYLQVHVARARIGHFPHKLYVPLSTFENNNTMFOILLGLKWKWFDLIGET 1219
 QY 419 EGVLEKPEPDI 430
 DB 1220 --VENEDPNNAFI 1229

RESULT 21
 HYSA_STRPN STANDARD; PRT; 1066 AA.
 ID_HYSA_STRPN
 AC Q54873; Q54874;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hyaluronate lyase precursor (EC 4.2.2.1) (Hyaluronidase) (Hase).
 GN Sp0314.
 OS Streptococcus pneumoniae.
 OC Streptococcus pneumoniae; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC BAA-334 / TIGR4;
 RX MEDLINE=21357209; PubMed=1163916;
 RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson M.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouri H., Wolf A.M., Ulteback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickenson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae.";
 RL Science 293:498-506(2001).
 RN [2]
 RP SEQUENCE OF 89-1066 FROM N.A.
 RC STRAIN-Type 23;
 RX MEDLINE=94156460; PubMed=8112843;
 RA Berry A.M., Lock R.A., Thomas S.M., Rajan D.P., Hansman D.,
 RA Paton J.C.;
 RT "Cloning and nucleotide sequence of the Streptococcus pneumoniae
 RT hyaluronidase gene and purification of the enzyme from recombinant
 RT Escherichia coli.";
 RL Infect. Immun. 62:1101-1108(1994).
 RN [3]
 RP CRYSTALLIZATION.
 RX MEDLINE=98234706; PubMed=9573623;
 RA Jedzejka M.J., Chantelat L., Mewbourne R.B.;
 RT "Crystallization and preliminary X-ray analysis of Streptococcus
 RT pneumoniae hyaluronate lyase.";
 RL J. Struct. Biol. 121:73-75(1998).
 CC -1- CATALYTIC ACTIVITY: Hyaluronate = N 3-(4-deoxy-beta-D-gluc-4-
 CC enuronosyl)-N-acetyl-D-glucosamine.
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan
 CC by an amide bond (Potential).
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE007344; AAK74491.1; -;
 DR EMBL; L20670; AAK53685.1; ALT INIT.
 DR EMBL; L20670; AAK53686.1; ALT_INIT.
 DR PIR; B95037; B95037.
 DR PDB; 1C82; 05-APR-01.
 DR PDB; 1F93; 16-MAY-01.
 DR PDB; 1LOH; 07-AUG-02.
 DR PDB; 1LXK; 07-AUG-02.
 DR PDB; 1NTN; 31-DEC-02.
 DR PDB; 1N7O; 31-DEC-02.
 DR PDB; 1N7P; 31-DEC-02.
 DR PDB; 1N7Q; 31-DEC-02.
 DR PDB; 1N7R; 31-DEC-02.
 DR TIGR; SP0314; -;
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR006192; LPXNG.
 DR InterPro; IPR003159; Lyase_8.
 DR InterPro; IPR004103; Lyase_8_C.
 DR Pfam; PF02018; CBM_4_9; 1.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF02378; Lyase_8_C; 1.
 DR Pfam; PF02884; Lyase_8_C; 1.
 DR TIGRfam; TIGR01167; LPXNG_anchor; 1.
 DR PROSITE; PS50847; GRAM_POS_ANCHORING; FALSE NEG.
 DR Lyase; Cell wall; peptidoglycan-anchor; signal; Complete proteome;

KW 3D-structure. 1 30-
 FT SIGNAL 31 1039 POTENTIAL.
 FT CHAIN 1040 1066 HYALURONATE LYASE.
 FT PROPEP 1036 1066 REMOVED BY SORTASE (POTENTIAL).
 FT SITE 1039 1039 LPXTG SORTING SIGNAL (POTENTIAL).
 FT MOD RES 1039 1039 AMIDE-LINKED TO CELL WALL (POTENTIAL).
 FT CONFLICT 108 108 H -> P (IN REF. 2).
 FT CONFLICT 115 108 L -> V (IN REF. 2).
 FT CONFLICT 139 139 I -> V (IN REF. 2).
 FT CONFLICT 211 211 L -> P (IN REF. 2).
 FT CONFLICT 236 236 S -> P (IN REF. 2).
 FT CONFLICT 290 290 A -> T (IN REF. 2).
 FT CONFLICT 313 313 E -> D (IN REF. 2).
 FT CONFLICT 340 340 T -> I (IN REF. 2).
 FT CONFLICT 613 613 C -> R (IN REF. 2).
 FT CONFLICT 658 658 P -> T (IN REF. 2).
 FT CONFLICT 821 821 G -> S (IN REF. 2).
 FT CONFLICT 848 848 V -> S (IN REF. 2).
 FT CONFLICT 853 853 F -> S (IN REF. 2).
 FT CONFLICT 907 907 R -> G (IN REF. 2).
 SQ SEQUENCE 1066 AA; 120771 MW; 81DB22A837BE61F9 CRC64;
 Query Match 4.7%; Score 105; DB 1; Length 1066;
 Best Local Similarity 21.2%; Pred. No. 19;
 Matches 96; Conservative 83; Mismatches 157; Indels 116; Gaps 26;
 8 PGKAFAGVYLVLEPIYDAVYATLASSRMHAYTPKSTLSKESRIKI-SSPOFAN---GE 62
 247 PDVASVYKNG--ILEPLKE-----GTTNVIVSKDGKVKIKIPLKILASVADAYTDRLDD 297
 63 WEYHISNT--EKPREVSRINPLEATFTIVLAYIOPTAEADLEIIIVSDPGYHSDGT 120
 298 WNGIILAGNQYVDSKNEQAKNLQELGKADSLSSISS-----QADRTYLMKFS 347
 121 ERTKTSNGEKTPLYSRAITEVEKTKGLSSAGLSVAVATSLSHRPVIVISNKGILNV 180
 348 NYKTSANITLVY---RLKEEMAK-----QVTNPSKRYODETVVTRVRDSMEWM 393
 161 AQIACVAYOKK--IGSGFDVAATVGLIVRRFOPALINDVFOVE---SDPE--KFPTE 233
 394 HK--HYINSEKSIYANMWD-----YELGTFRALINNTLSLMEKFSDEIKKYTDV 441
 234 LKKLIESNWEKEKRCCTLPYGIKLMGDVKGSGSETPKLVSRVLQWKKEPSSVYVDQL 293
 442 IEKFP---DPEHFRKTTDNPFKALGKLVDMGRV-KYIAGL--RKDDQELISSTI---- 491
 294 NSANTQFMKEKLEMEKTKDSPEYIKELDHSVEPLVTAIKNIR-KGIQAL---TQKSEV 349
 492 --RSIEOVFKLVDDQEGFYOD-GSYI---DHTNVAVTGAYGVNLIDGLSQLPVIQKTKN 545
 350 PIEPD-----VOTQLDRCGEIRGCVGVVPGA---GGYDAIAVL--- 386
 546 PIDKDKMOTMTHWIDKSRAPLLVNGELMDMS-----GRSISRANSGHAAVVEVLKGI 599
 387 --VLENOVGNFKQ-----KTLLENPDYFRNYY 410
 600 HRIADMSEGETKQCLQSLVKTIQVSDSYVDV 631
 RESULT 22
 VPS_BTVA2A
 ID_VPS_BTVA2A STANDARD; PRT; 526 AA.
 AC P30209;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Outer capsid protein VPS.
 GN 56.
 OS Bluetongue virus (serotype 2 / isolate USA).
 OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
 OX NCBI_TaxID=10907;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE-90209358; PubMed-2157314;
RA Hirasawa T., Roy P.;
RT "The complete nucleotide sequence of VP5 of a strain of bluetongue
RT virus of serotype 2 isolated in the USA reveals its close
RT relationship with a virus of serotype 1 isolated in Australia.";
RL Virus Res. 15:107-112(1990)
CC -1- FUNCTION: THE VP5 PROTEIN IS ONE OF THE TWO PROTEINS (WITH VP2)
CC -1- WHICH CONSTITUTE THE VIRUS PARTICLE OUTER CAPSID.
CC -1- SIMILARITY: BELONGS TO THE REOVIRUSES VP5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X62283; CAA44172.1; -
CC DR PIR; A43486; P5XRBU.
CC DR InterPro; IPR000145; Orbi VP5.
CC DR Pfam; PF00901; Orbi_VP5; I.
CC KW Coat protein
SQ SEQUENCE 526 AA; 58953 MW; DE50D6013B983A04 CRC64;

Query Match 4.7%; Score 104.5; DB 1; Length 526;
Best Local Similarity 23.7%; Pred. No. 7.5;
Matches 64; Conservative 46; Mismatches 117; Indels 43; Gaps 12;

QY 162 LSHF--IPNVISTN-KDILHNVAQIAHCYAKKIGSGFD--VATAYGLIVRRF-- 211
DB 8 LSRFKKVGSAALTSNAKAKIYVSTICKAERFAEISGASATDGLVQSGHSLTSESYGE 67
QY 212 ---OPALIN-----DVFQVLESDEKPETELKLIENSWEKHERCTLPVIGIKLMDV 262
DB 68 SVKQAVLNVLAVGSGEIPDPSPGQGMQTKLKELE--EGRNELVRLKYNDKI--KE 121
QY 263 KGSSETPKLVSRVLQMKKE--PESSVYVDQLNSANLQFKELRENRKEDYSDPEYTI 319
DB 122 KFGSELQGVYEFMNAKEEVAEAKQFDLSKAVNSYKILTEEOQQRRL-----ALMLQ 177
QY 320 KELDSVEPLVYAINIKRQALQALQKSEVPPIEDPVQQLDRCCIEFGCVGV----- 373
DB 178 KEIGERTVEYFAMVKEYENRKIDAL--KSAIEIERDMQES--EAIQEIAGMTADVLEAASE 233
QY 374 -VPGAGYDAIVAVLVLENQVGNFKQKTLEN 402
DB 234 EYPLIGAGMATVATGTAIEGAYKLKAVIN 263

RESULT 23
DYN1_RAT STANDARD; PRT; 851 AA.
AC P21575;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dynamitin-1 (Ec 3.6.1.50) (D100) (Dynamitin, brain) (B-dynamitin).
CN DNM1 OR DNM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=90384564; PubMed=2144893;
RA Obar R.A., Collins C.A., Hammarback J.A., Shpetner H.S.,
RA Vallee R.B.;
RT "Molecular cloning of the microtubule-associated mechanochemical
RT enzyme dynamitin reveals homology with a new family of GTP-binding
RT proteins.";
RL Nature 347:256-261(1990).

CC -1- FUNCTION: MICROTUBULE-ASSOCIATED FORCE-PRODUCING PROTEIN INVOLVED
CC IN PRODUCING MICROTUBULE BUNDLES AND ABLE TO BIND AND HYDROLYZE
CC GTP. MOST PROBABLY INVOLVED IN VESICULAR TRAFFICKING PROCESSES, IN
CC PARTICULAR ENDOCYTOSIS.
CC -1- CATALYTIC ACTIVITY: GTP + H(2)O = GPP + phosphate.
CC -1- SUBCELLULAR LOCATION: MICROTUBULE-ASSOCIATED.
CC -1- TISSUE SPECIFICITY: CONFINED TO THE BRAIN.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN NEURONS AFTER MATURATION.
CC -1- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.
CC -1- SIMILARITY: Contains 1 PH domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X54531; CAA38397.1; -
CC DR PIR; S11508; S11508.
CC DR HSSP; 005193; IDYN.
CC DR InterPro; IPR001401; Dynamitin.
CC DR InterPro; IPR000375; Dynamitin_central.
CC DR InterPro; IPR003130; GBD.
CC DR InterPro; IPR001849; PH.
CC DR Pfam; PF00350; dynamitin_1.
CC DR Pfam; PF01031; dynamitin_2; 1.
CC DR Pfam; PF02212; GBD; 1.
CC DR Pfam; PF00169; PH; 1.
CC DR PRINTS; PR00195; DYNAMIN.
CC DR SMART; SMO0053; DYNC; 1.
CC DR SMART; SMO0233; PH; 1.
CC DR PROSITE; PS00410; DYNAMIN; 1.
CC DR PROSITE; PS50003; PH DOMAIN; 1.
CC KW Hydrolyase; Motor protein; GTP-binding; Microtubules; Multigene family;
KW Endocytosis.
FT NP_BIND 38 45 GTP (POTENTIAL).
FT NP_BIND 136 140 GTP (POTENTIAL).
FT NP_BIND 205 208 GTP (POTENTIAL).
FT FT DOMAIN 519 625 PH.
FT FT DOMAIN 753 851 PRO-RICH.
SQ SEQUENCE 851 AA; 95927 MW; 5D9294C5D1015A88 CRC64;

Query Match 4.7%; Score 104.5; DB 1; Length 851;
Best Local Similarity 22.6%; Pred. No. 15;
Matches 67; Conservative 49; Mismatches 113; Indels 67; Gaps 13;

QY 76 EVQSRINPFLATIFIVLAVIQTEAPFLDIIVSD-----PGY-----HSQEPTEKTS 125
DB 188 KIAKEVDPOGORTIGVITKLDLMDEGTDARDVLENKLLPRRGYIGVYVNSQKQIDGKD 247
QY 126 SNG-----EKTFLHSRAITEVEKTGL-----GSSAGLVSVVATSLLS- 163
DB 248 ITALAALAEKRFLESHPSYRLADMGTPYQKLVNQQLTNHRIPTLGLARKLQSQLSLI 307
QY 164 -----HPIP-NVISTNKDILHNVAQIAHCYAKKIGSGFDVATAYGLIVRRRPPA 214
DB 308 EKEVDEYKNFPPDPARTKALQLQVOQFADPEKRIEGSDQIDT-----YELSGA 360
QY 215 LINDVFOVLSDPKFPTLKLIESNWEKHERCTLPYGIKLMGDVKGSETPKLVSR 274
DB 361 RINRIH-----RPFPELVKM--EPDEKELRREISYAIKNIHG--IRTLFPTDLAFE 410
QY 275 VLQWK-KEDESSV-----VYDQLNSANLQFMKLEB--MRKDYSDPEYTIKE 321
DB 411 ATVKKQYQVKKEPSIKCDVWVSELSTIRKCSKQLQQYPLRERMERIVTTTHIRE 466

RESULT 24
DYN1_MOUSE STANDARD; PRT; 861 AA.

Oy	164	-----HFIF-NVISTNKOILHNVAQIACVAKOKGSGFDVATAYGLIVRRFOPA	214
Dd	308	EKEVDENKFNFRDDDPARKTKALLQNVQAFADPEFKRIEGSSDQIDT-----YELSGA	360
Oy	215	LINDVEQVLSPDEKPFTELKKLIESNWEEKHRECTLPGYIKILMGDKVGSGSTPKLVSR	274
Dd	361	RINRIPH-----ERFPFELVGM---EFDEKELRREISALKNIHG--IRGLFLTPDMAFE	410
Oy	275	VLOMKEREES-----SVYYDLQNSANLPMELE--MRKYSDEPTYIKE	321
Dd	411	TIVKQVKRKIREPCCKCVMWVISLELSITVRCTCKKLQQPYRLREMERIVTTHIRE	466
<hr/>			
RESULT 25			
RESA_PLAFF	STANDARD;	PRT.	1073 AA.
ID	RESA_PLAFF		
AC	P18330;		
DT	01-JAN-1990 (Rel. 13, Created)		
DT	01-JAN-1990 (Rel. 13, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DN	Ring-infected erythrocyte surface antigen precursor.		
GN	RESA.		
OS	Plasmodium falciparum [isolate FC27 / Papua New Guinea].		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NDBI_TaxId=5837;		
RN	[1]		
RX	SEQUENCE FROM N.A.		
RA	MEDLINE=6706671.0; PubMed=3537955;		
RA	Favaloro J.M., Coppel R.L., Corcoran L.M., Foote S.J., Brown G.V.,		
RA	Andera R.F., Kemp D.J.;		
RT	"Structure of the RESA gene of Plasmodium falciparum.";		
RL	Nucleic Acids Res. 14:8265-8277(1986).		
CC	-1- FUNCTION: RESA MAY DISRUPT THE NORMAL INTERMOLECULAR INTERACTIONS OF THE CYTOPLASMIC DOMAIN OF BAND 3 AND THEREBY FACILITATE THE INVASION OF THE RED CELL MEMBRANE WHICH IS NECESSARY FOR THE FORMATION OF THE PARASITOPOROUS VACCINE.		
CC	-1- SUBCELLULAR LOCATION: PROBABLY LOCATED ON THE CYTOPLASMIC FACE OF THE MEMBRANE WHERE IT ASSOCIATES WITH COMPONENTS OF THE MEMBRANE SKELETON.		
CC	-1- PTM: THE TYR RESIDUES IN THE VARIANT TETRAMERIC SEQUENCES IN THE RESA REPEAT ARE POSSIBLY PHOSPHORYLATED (BY HOMOLOGY WITH BAND 3).		
CC	-1- SIMILARITY: THE N-TERMINAL SEQUENCE OF BAND 3 SHOW HOMOLOGY WITH THE REPEAT SEQUENCES OF RESA.		
CC	-1- SIMILARITY: Contains 1 J domain.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See http://www.isb-sib.ch/announce or send an email to licenses@isb-sib.ch).		
CC	-----		
DR	EMBL; X04517; CAA28241.1.; -		
DR	EMBL; X05182; CAA28817.1.; -		
DR	EMBL; A00661; CAA00077.1.; -		
DR	PIR; A25526; A25526.		
DR	HSSP; P08622; IBOZ.		
DR	InterPro; IPR001623; DnaJ N.		
DR	InterPro; IPR003095; Hsp_DnaJ.		
DR	Pfam; PF00226; DnaJ_1.		
DR	PRINTS; PR00625; DNADJPROTEIN.		
DR	SMART; SM00271; DnaJ_1.		
DR	PROSITE; PS00636; DNADJ_1; 1.		
DR	PROSITE; PS50076; DNADJ_2; 1.		
KW	Malaria; Antigen; Glycoprotein; Repeat; Signal.		
KW	SIGNAL		
FT	CHAIN	1	65
FT		66	1073
FT			
FT	DOMAIN	436	504
FT		521	589
FT	DOMAIN	891	1073
FT		71	
FT	CARBOHYD		
<hr/>			
RING-INFECTED ERYTHROCYTE SURFACE ANTIGEN. TANDDEM REPEATS 1. J-DOMAIN TANDDEM REPEATS 2. N-LINKED (GLCNAC. .) (POTENTIAL).			

FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 773 773 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 777 777 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 378 378 S -> T (IN NP7).
 SQ SEQUENCE 1073 AA; 124906 MW; A4D3C37E10BA6D90 CRC64;
 Query Match 4.7%; Score 104.5; DB 1; Length 1073;
 Best Local Similarity 18.8%; Pred. No. 21;
 Matches 86; Conservative 87; Mismatches 175; Indels 109; Gaps 21;
 QY 19 VLEPIYDAVATLSSRMNAVITPKGSLKESRIKISSPPANGEMEY-----HISNTE 72
 DB 666 ILQP-----CIAGDSKMNVPITTKLEGLKSRFDI--PLLESLRMTFKVATHTKKSKS 718
 QY 73 KREVOQR-----INPLEATIFIVLAVIOPTFADEIITISDPGHSQDET 122
 DB 719 SAKKIQORTQANKQELANINNNLMSTL---KEYLSSSE--QMNSTTYNEINNSVNDGN 773
 QY 123 KTSNGEKTFLYHSR-----ATVEKGTGLSSAGLVSVVATSLSHFIPNV 170
 DB 774 QSKNISDLSYTDQKEILEKIVSYVIDISLYDIENALNAAEQLLSDNSVD-----EKT 827
 QY 171 STNNKILHNVAGIACFYA---QKIGSGFDVATAIYGLIVRRPOPALINDVFOVLSD 226
 DB 828 KRAOSLKKLSSIMERYAGKKNKKKKNFTKD--IVGYIMG-----ISTINTEMNQ 880
 QY 227 PEKPFTELKTLIESNWEKHERCTLPYGIKLMGVKSGSEPKLYSRVLQMKKEPRES 286
 DB 861 NENVEHVOHNAEENVEHDAE-----NVEHDAE--ENVEHDAEENVEHDAE 926
 QY 287 SVVYQOLNSA--NIQFMKELRMRREKYSDPEYIKELDSVEPLTVAIKIRKGLQALT 344
 DB 927 NVEHDAEENVEENVEEVE--NVEENVEENVEENVEENVEENVEENVEENVEENVE 978
 QY 345 QKS--EVPLEPVDQTOGLDRCOEIPGCVGVVPGAGYDAIAY--LLENOGVKFKXTLEN 402
 DB 979 ENVEENVEENVEENVEEVEE-----YEEENVEEENVEENVEENVEENVEEN 1024
 QY 403 PDYFNHYVWD-----LEEQTGVLEKPEPDY 429
 DB 1025 VE--ENVEEVEENVEENVEENVEENVEENVEENVEENVEEY 1059
 RESULT 26
 GRSB_BACBR STANDARD; PRT; 4451 AA.
 ID_GRSB_BACBR STANDARD; PRT; 4451 AA.
 AC P14688;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Gramicidin S synthetase II [includes: ATP-dependent proline adenylyase
 DE (Valine activase); ATP-dependent ornithine adenylyase (OrnA) (Ornithine
 DE activase); ATP-dependent leucine adenylyase (LeuA) (Leucine activase)].
 GN GRSB OR GRS2.
 OS *Bacillus brevis* (Brevibacillus brevis).
 OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
 CC NCBI_TaxID=1393;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 9999 / IFO 15520 / JCM 8504 / NCTC 7096;
 RX MEDLINE=9221998; PubMed=1560782;
 RA Turgay K., Krause M., Marahel M.A.;
 RT "Four homologous domains in the primary structure of GrsB are related
 RT to domains in a superfamily of adenylylate-forming enzymes.";
 RL Mol. Microbiol. 6:529-546 (1992).
 RN [2]
 RP SEQUENCE OF 1-948 FROM N.A.
 RC STRAIN=Nagano;
 RX MEDLINE=92041751; PubMed=1939016;
 RA Hori K., Yamamoto Y., Tokita K., Saito F., Kurotsu T., Kanda M.,
 RA Okamura K., Furuyama J., Saito Y.;
 RT "The nucleotide sequence for a proline-activating domain of

RT gramicidin S synthetase 2 gene from *Bacillus brevis*.";
 RL J. Biochem. 110:111-119 (1991).
 RN [3]
 RP SEQUENCE OF 1-143 FROM N.A.
 RC STRAIN=ATCC 9999 / IFO 15520 / JCM 8504 / NCTC 7096;
 RX MEDLINE=9008776; PubMed=2477357;
 RA Kraetzschmar J., Krause M., Marahel M.A.;
 RT "Gramicidin S biosynthesis operon containing the structural genes
 RT grsA and grsB has an open reading frame encoding a protein homologous
 RT to fatty acid thioesterases.";
 RL J. Bacteriol. 171:5422-5429 (1989).
 RN [4]
 RP SEQUENCE OF 1-15, AND CHARACTERIZATION.
 RC STRAIN=Nagano;
 RX MEDLINE=92011463; PubMed=1917901;
 RA Kurotsu T., Hori K., Kanda M., Saito Y.;
 RT "Characterization and location of the L-proline activating fragment
 RT from the multifunctional gramicidin S synthetase 2.";
 RL J. Biochem. 109:763-769 (1991).
 CC -1- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME, ABLE TO
 CC ACTIVATE AND POLYMERIZE THE AMINO ACIDS PRO, VAL, ORN AND LEU.
 CC ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
 CC -1- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTHETINES.
 CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
 CC ANTI-BIOTIC GRAMICIDIN S (D-PHE-PRO-VAL-ORN-LEU) 2.
 CC -1- SUBUNIT: LARGE MULTIZYME COMPLEX OF GRsA AND GRsB.
 CC -1- DOMAIN: CONSISTS OF FOUR MODULES, AND HARBOURS A PUTATIVE
 CC THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE
 CC INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE
 CC FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE
 CC ADENSYLATION, THIOYLATION, CONDENSATION (NOT FOR INITIATION
 CC MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION
 CC (OPTIONAL).
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -1- SIMILARITY: Contains 4 acyl carrier domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 CC EMBL, X61658, CAA43838.1; -;
 CC EMBL, M29703, AA58719.1; -;
 CC EMBL, X15577, CAA33604.1; -;
 CC EMBL, D00938, BAA00778.1; -;
 CC PIR, S20542; YGSGS2.
 CC HSSP, P14687; 1AMU.
 CC InterPro: IPR000873; AMP-bind.
 CC InterPro: IPR001242; Condensatn.
 CC InterPro: IPR006163; Pp bind.
 CC InterPro: IPR006162; Pantane atach.
 CC InterPro: IPR000379; Ser estse site.
 CC InterPro: IPR001031; Thioesterase.
 CC Pfam, PF00501; AMP-binding; 4.
 CC Pfam, PF00668; Condensation; 4.
 CC Pfam, PF00550; Pp-binding; 1.
 CC Pfam, PF00975; Thioesterase; 4.
 CC PROSITE, PS00612; PHOSPHOPANTHETINE, 4.
 CC PROSITE, PS00455; AMP BINDING; 4.
 CC PROSITE, PS50075; ACP_DOMAIN; 4.
 CC LIGase; Antibiotic biosynthesis; Multifunctional enzyme;
 CC Repeat; Phosphopantetheine; Hydrolase.
 FT INIT MET 0
 FT REPEAT 466 1043 DOMAIN 1 (PROLINE-ACTIVATING).
 FT REPEAT 1521 2080 DOMAIN 2 (VALINE-ACTIVATING).
 FT REPEAT 2538 3135 DOMAIN 3 (ORNITHINE-ACTIVATING).
 FT REPEAT 3591 4173 DOMAIN 4 (LEUCINE-ACTIVATING).
 FT DOMAIN 975 1042 ACYL CARRIER (ACP) 1.
 FT DOMAIN 2011 2078 ACYL CARRIER (ACP) 2.

FT DOMAIN 3057 3124 ACYL CARRIER (ACP) 3.
 FT DOMAIN 4095 4162 ACYL CARRIER (ACP) 4.
 FT BINDING 1005 1005 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT BINDING 2041 2041 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT BINDING 3087 3087 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT BINDING 4125 4125 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT CONFLICT 274 274 H -> D (IN REF. 2).
 FT CONFLICT 418 418 K -> R (IN REF. 2).
 FT CONFLICT 654 664 NAVICUTKFF -> TCSDFVCOEL (IN REF. 2).
 FT CONFLICT 941 946 HVALH -> QLPUP (IN REF. 2).
 SQ SEQUENCE. 4451 AA; 510036 MM; E0029C9B51F5A4B7 CRC64;

Query Match 4.7%; Score 104.5; DB 1; Length 4451;
 Best Local Similarity 21.6%; Pred. No. 1.6e+02;
 Matches 76; Conservative 48; Mismatches 139; Indels 89; Gaps 17;

QY 36 HAVITPKGTSLKESRIKISSPOFANGEMEHYSSNTEKREVOGRINPLEATITIVL-- 93
 DB 3605 HAIWEGQALTYHEINIKANOLARLREKGVTPN--HVALWTERSLMTIVGIFSLKA 3662
 QY 94 --AYIQTEAFLEIITISDPGHSOEDTETKSSNGEKFVLSRAITEV----- 142
 DB 3663 GGAIVP-----IDPAT--PQERQVLEDSGALLLQSHVANKLPVDIEMDL 3709
 QY 143 -----EKTGL--GSSAGIVSVATSLSHFIPNVISTNKDILAHVAQIAHCYAOKK 191
 DB 3710 TDEQNVVEDGTNLPMNOSTDLAYIITYS-----GTTGKPGVWIEHOSIINCLOMKR 3762
 QY 192 ----IGSGDVATATYGL----IVRRFOPALINDVFOVLESDEKPEPELKLISNME 243
 DB 3763 EEEYEPFG--DTALQVSPAFDFVSLFAPILAGATSVLPKEEADPVALKXKLAS-- 3818
 QY 244 EKHERCTLPYGIKILMG---DYK-----GSEETP--KLVSRLQWKK----- 261
 DB 3819 ---EETTHYYGVPSLSALIDWSSKDLNRCVTYLGEEKLPQIYKXKKEKKKELEVNN 3875
 QY 282 --KPESSVYVDOLNANLQFMKELREMEKXDSDEPTYIKELDSVEPLTV 331
 DB 3876 EKGPTENSVVTTIMR--DIQVEQETITGRPL--SNVDYIVVNCNQLQGVGV 3923

RESULT 27
 PAFPA CAVPO STANDARD; PRT; 436 AA.
 ID PAFPA CAVPO STANDARD; PRT; 436 AA.
 AC P70683;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Platelet-activating factor acetylhydrolase precursor (BC 3.1.1.47)
 DE (PAF acetylhydrolase) (PAF 2-acetylhydrolase) (LDL-associated
 DE phospholipase A2) (LDL-PLA(2)) (2-acetyl-1-alkylglycerophosphocholine
 DE esterase) (1-alkyl-2-acetyl-1-glycerophosphocholine esterase).
 GN PLA2G7 OR PAFPAH.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.
 OX NCBI_Taxid=10141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Hartley; TISSUE=Liver;
 RX MEDLINE=97103479; PubMed=8947850;
 RA Karsawa K., Kuge O., Kawasaki K., Nishijima M., Nakano Y., Tomita M.,
 RA Yokoyama K., Setaka M., Nojima S.;
 RT "Cloning, expression and characterization of plasma
 RT platelet-activating factor-acetylhydrolase from guinea pig";
 RL J. Biochem. 120:838-844(1996).
 CC -1- FUNCTION: MODULATES THE ACTION OF PLATELET-ACTIVATING FACTOR (PAF)
 CC BY HYDROLYZING THE SN-2 ESTER BOND TO YIELD THE BIOLOGICALLY
 CC INACTIVE LYSO-PAF. HAS A SPECIFICITY FOR SUBSTRATES WITH A SHORT
 CC RESIDUE AT THE SN-2 POSITION. IT IS INACTIVE AGAINST LONG-CHAIN
 CC PHOSPHOLIPIDS.
 CC -1- CATALYTIC ACTIVITY: 2-acetyl-1-alkyl-sn-glycero-3-phosphocholine +
 CC H(2)O = 1-alkyl-sn-glycero-3-phosphocholine + acetate.

CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: Plasma.
 CC -1- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: D67037; BAA11054.1; -
 CC PIR: JCS021; JCS021.
 CC InterPro: IPR000734; Lipase.
 CC InterPro: IPR005065; PAF-AH_p_II.
 CC InterPro: IPR000379; Ser_estre_sic.
 CC Pfam: PF03403; PAF-AH_p_II; 1.
 CC PROSITE: PS00120; LIPASE_SSR; 1.
 CC KMW Hydrolase; Lipid degradation; Glycoprotein; signal.
 CC FT SIGNAL 1 21
 CC CHAIN 22 436
 CC FT ACT SITE 271 271
 CC FT ACT SITE 294 294
 CC FT ACT SITE 349 349
 CC FT CARBOHYD 76 76
 CC FT CARBOHYD 200 200
 CC FT CARBOHYD 324 324
 CC SQ SEQUENCE 436 AA; 49062 MM; C359D96E392FPE11 CRC64;
 Query Match 4.6%; Score 103.5; DB 1; Length 436;
 Best Local Similarity 22.8%; Pred. No. 6.8;
 Matches 71; Conservative 45; Mismatches 118; Indels 77; Gaps 14;

QY 62 EWEHISSNTEK--REVOSRINPLEATITIVLAYIOTPEAFLEIITIS 110
 DB 133 KNSPLKTGKEKPLIFSHGLCAFRISYSAIGIELASHGFIYAAVEHDES-----A 184
 QY 111 DPGHSOEDTETKSSNGEKFVLYH--SRATTEVKTGLSSAGIVSVVATSLS--H 164
 DB 185 AATYTFQ--DAPAESGNRSMIYKGNLEERKQLRGRESCQALSIDBGE 241
 QY 165 FIPNVISTNKDILAHVAQIAHCYAOKK--IGSGDVATATYGLIVRRFOPALIND--V 219
 DB 242 FVAVNLDLNPDI---QDLKSLDRSKVAIIIGSHSGGATVYQTLSEDRFRGIALDPM 297
 QY 220 FOVLESDEKEPTEL-----KKLIESNMEKHERCTLPYGIKILMGDYKGS 266
 DB 298 FVVGEDVSKIPQPLPFINSEYFQSANDTKLEKRYQPKER-----KMIA--VKGSV 348
 QY 267 E-----TPKIVSRVLQMKKEPPESSVYVDOLNANLQFMKELREMEKXP-- 312
 DB 349 HHNFVDFTFATGKIIGMLSLKGR--IDSEVAMDLINFAFLQKYIGLDKNFQNSL 406
 QY 313 --SDPEYIKS 321
 DB 407 MEGDENLIPR 417

RESULT 28
 MUTS BACSU STANDARD; PRT; 858 AA.
 ID MUTS BACSU STANDARD; PRT; 858 AA.
 AC P49849;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA mismatch repair protein muts.
 GN MUTS.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_Taxid=1423;
 RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96349107; PubMed=8760914;
RA Ginetti F., Perego M., Albertini A.M., Gallizi A.;
RT "Bacillus subtilis muts muts operon: identification, nucleotide
sequence and mutagenesis."
RL Microbiology 142:2021-2029(1996).
RM [2]
RN SEQUENCE FROM N.A.
RP STRAIN=168;
RC MEDLINE=98044033; PubMed=9384377;
RX Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RB Borriis R., Boutsier L., Brans A., Braun M., Briganti S.C., Bron S.,
RC Brouillet S., Brunsch C.V., Caldwell B., Capuano V., Carter N.M.,
RD Choi S.K., Codan J.J., Comerton I.F., Cummings N.J., Daniel R.A.,
RE Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RF Entian K.D., Errington J., Fadrer C., Ferrari E., Foulger D.,
RG Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi A., Galleron N.,
RH Ghm S.Y., Glaser P., Goffeau A., Goldightly E.J., Grandi G.,
RI Giuseppe G., Guy B.J., Haga K., Hahsch J., Harwood C.R., Henaut A.,
RJ Hilbert H., Holsappel S., Hosono S., Hullio M.F., Itaya M., Jones L.,
RK Joris B., Karamata D., Kasahara Y., Kleaer-Blanchard M., Klein C.,
RL Kobayashi Y., Koester P., Koningsbein G., Krogh S., Kumano M.,
RM Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RN Lee S.M., Levine A., Liu H., Masuda S., Manel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RB Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RC Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RD Pirescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RE Rieger M., Rivola C., Rochna E., Roche B., Rose M., Sadate Y.,
RF Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RG Seliguchi J., Sekowska A., Seror S.J., Serior P., Shin B.S., Soldo B.,
RH Socolnik A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RI Taneuchi M., Tanakoshi A., Tanaka T., Terperera P., Tognoni A.,
RJ Tosato V., Uchiyama S., Vandembol M., Vannier F., Vassarotti A.,
RK Viari A., Wandut R., Wedler E., Wedler H., Weltzenegger T.,
RL Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumschein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis."
RL Nature 390:249-256(1997).
RL Nature 390:249-256(1997).
CC -1- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is possible that it carries out the mismatch recognition
CC step. This protein has a weak ATPase activity (By similarity).
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.ibd-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U27343; AAB19235.1; ALT_INIT.
CC EMBL; Z99112; CAB13577.1; -
CC PIR; C69663; C69663.
CC Subtilase; Bcl1403; muts.
DR HAVAP; MF_00096; -; 1.
DR InterPro; IPR005748; MutS1.
DR InterPro; IPR000432; Muts_C.
DR InterPro; IPR002863; Muts_N.
DR Pfam; PF01624; Muts_I; 1.
DR Pfam; PF05188; Muts_II; 1.
DR Pfam; PF05192; Muts_III; 1.
DR Pfam; PF05190; Muts_IV; 1.
DR Pfam; PF00488; Muts_V; 1.
DR ProDom; PD001263; Muts_C; 1.
DR SMART; SMO0534; MutsSac_1.
DR SMART; SMO0533; Mutsd; 1.
DR TIGRFAMs; TIGR01070; muts1; 1.
DR PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.

FW	DNA repair; ATP-binding; DNA-binding; Complete proteome.
FT	NP_BIND 602 609 ATP (POTENTIAL).
SO	SEQUENCE 858 AA; 97590 MW; 641BDD6BCA4BC5D CRC64;
Query Match	4.6%; Score 103.5; DB 1; Length 858;
Best Local Similarity	20.6%; Pred. No. 18;
Matches	91; Conservative 69; Mismatches 134; Indels 147; Gaps 22
QY	1 MSKAFAPGKAFLAGYLVLEPIYDA-VYVALSSHMAVITPKGTSIKSRKISDPOFA 59
DB	161 ISIIYSVGAREIIVSGSL-----DADIVQAQLREKCGATIS-----IE 197
QY	60 NGEWMEYHIS-----SNTKEPKEVQSSINPLEATITFYALVQPTPEADLE---IIITYSD 111
DB	198 DGETDEHVTIIEHLNNEIDITKTETFLVLYLLYKTKQKRSIDHQLQVVELEAKMKIDLYSK 257
QY	112 PGVHSGEDIEFTKTSNGEKTPLVHSRAITVEKTKLS-----SAGIVSV----- 156
DB	258 ---RNLELLEITRKRKKKSGSLM-----LIDETKTPMGRLKKQMTDRLPLRVNOIEROE 310
QY	157 VATSLLSHPF-----NVISTN-----KQILHN-----VA 187
DB	311 MVEETLASHFFEREDLRERLKEVYDLERLAGRVAGGNVARDLIQKLSKQVPEIKQLVA 370
QY	182 QIAHCAQCKKISSGPDVATVATVIGLIVRRFPQALINDVFOVLESD-PEKFPTELK--LI 238
DB	371 SLAHDAKERA-----KRIDPC--GDVLELLEALYENPLSVKGNLI 412
QY	239 ESNWEKHEKRCITLPYGIKLMGDVKGSETPK-LVSRVLQWKKEKPESSV----- 288
DB	413 KQGIYNCK-----LDEYDASRNGKQMTIARLEQOEREYTGIRLSKGFNKVPG 455
QY	289 VYDQLSANLQEMKELREMKREKYDSDEPTYI-KEILDHSEVEPLVAINIKRIGLQALTKS 347
DB	460 YYIEVYKVALHLLBESRYERNETLNNARYITTPELKEKCALILAEANNI----- 508
QY	348 EVPIEPDVOTQLDRC-QEIP 367
DB	509 -CELEVELEFTELRERKVKQYIP 528
RESULT 29	
SSM4_SCHPO	STANDARD; PRT; 670 AA.
AC	042667;
DT	28-FEB-2003 (Rel. 41, Created)
DT	28-FEB-2003 (Rel. 41, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	Microtubule associated protein smu4.
GN	SSM4 OR SPAC27D7.13C OR SPAC637.01C.
OS	Schizosaccharomyces pombe (Fission Yeast).
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;
OX	NCBI_TaxID=4896;
RN	1]
RP	SEQUENCE FROM N.A.
RC	MEDLINE=97311255; PubMed=9167972;
RA	Yamashita A., Matanabe Y., Yamamoto M.,
RA	"Microtubule-associated coiled-coil protein Smu4 is involved in the
RA	meiotic development in fission yeast.";
RA	Genes Cells 2:155-166(1997).
RA	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=972;
RA	MEDLINE=21848401; PubMed=11859360;
RA	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA	Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA	Collins K., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA	James K., Jones I., Jones W., Leather S., McDonald S., McLean J.,

148 TYRDKGL-SKRF-----EKKLLNNKKNKINFKTKKEDFVLMKMSKXNPSWMSK 201
 QY 246 -----HERCTLPYGIKLL--MGDYKGSSET---PGLVSRVLQ----- 277
 Db 202 HRRPGMHIECS-AISDKLTGNI FDIHGSSDLIFPHHEHETIOSKANSKFNWNTWHTG 260
 QY 278 ---WKEKEESSVYVDQNSANLQFMKELAREKREKSDPETYIKELDSVEPLTVAIK 334
 Db 261 WMFKEKESKKS-----LGNCL-----IREIVKDYADLTLYFLTSHRSQITYSVD 309
 QY 335 NIKRGLQALTKC-----SEVPIDEDVOTQLDRQCF---IPGCVGVVPPAGAGVD 381
 Db 310 NNEKKSAAVKRLYRSLHNTYFNPLPKKRTSKTIDANNDDENTP-----K 356
 QY 382 AIAVLV-LENOVGNFKOTLENDYFHNYYVDLEBOTE--GVLEKPEPDI 430
 Db 357 ASVLPFSISROINYNFKNKPKEKANQSLV---LKNLSNITIGLINENPDSL 404

RESULT 31
 ID DYHC CAEBL STANDARD; PRT; 4568 AA.
 AC 019020:
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Dynein heavy chain, cytosolic (DYHC).
 GN DHC-1 OR T21E12.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=96114101; PubMed=8674131;
 RA Lye R.J., Wilson R.K., Waterston R.H.;
 RT "Genomic structure of a cytoplasmic dynein heavy chain gene from the
 RT nematode Caenorhabditis elegans."
 RL Cell Motil. Cytoskeleton 32:26-36(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Du Z., Maggi L.;
 RN Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
 RP [3]
 RP REVISIONS.
 RA Waterston R.;
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: Cytoplasmic dynein acts as a motor for the intracellular
 CC retrograde motility of vesicles and organelles along microtubules.
 CC Dynein has ATPase activity; the force-producing power stroke is
 CC thought to occur on release of ADP.
 CC -1- SUBUNIT: Consists of at least two heavy chains and a number of
 CC intermediate and light chains.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- DOMAIN: Dynein heavy chains probably consist of an N-terminal stem
 CC (which binds cargo and interacts with other dynein components),
 CC and the head or motor domain. The motor contains six tandemly-
 CC linked AAA domains in the head, which form a ring. A stalk-like
 CC structure (formed by two of the coiled coil domains) protrudes
 CC between AAA 4 and AAA 5 and terminates in a microtubule-binding
 CC site. A seventh domain may also contribute to this ring; it is not
 CC clear whether the N-terminus or the C-terminus forms this extra
 CC domain. There are four well-conserved and two non-conserved ATPase
 CC sites, one per AAA domain. Probably only one of these (within AAA
 CC 1) actually hydrolyzes ATP, the others may serve a regulatory
 CC function.
 CC -1- SIMILARITY: Belongs to the dynein heavy chain family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL, L33260; AAC37251.1; -
 DR EMBL, U80440; AAK21472.1; -
 DR WormRep; T21E12.4; CE23997.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR004273; Dynein_heavy.
 DR Pfam; PF03028; Dynein_heavy; 1.
 DR SMART; SM00382; AAA; 4.
 KW Motor protein; Dynein; Microtubules; ATP-binding; Repeat; Coiled coil.

FT DOMAIN 1 1826 STEM (BY SIMILARITY).
 FT DOMAIN 1827 2049 AAA 1 (BY SIMILARITY).
 FT DOMAIN 2118 2394 AAA 2 (BY SIMILARITY).
 FT DOMAIN 2498 2747 AAA 3 (BY SIMILARITY).
 FT DOMAIN 2842 3111 AAA 4 (BY SIMILARITY).
 FT DOMAIN 3132 3432 STALK (BY SIMILARITY).
 FT DOMAIN 3496 3725 AAA 5 (BY SIMILARITY).
 FT DOMAIN 3954 4169 AAA 6 (BY SIMILARITY).
 FT DOMAIN 587 652 COILED COIL (POTENTIAL).
 FT DOMAIN 814 844 COILED COIL (POTENTIAL).
 FT DOMAIN 1241 1274 COILED COIL (POTENTIAL).
 FT DOMAIN 1324 1340 COILED COIL (POTENTIAL).
 FT DOMAIN 1559 1591 COILED COIL (POTENTIAL).
 FT DOMAIN 3132 3329 COILED COIL (POTENTIAL).
 FT DOMAIN 3339 3432 COILED COIL (POTENTIAL).
 FT DOMAIN 3707 3739 COILED COIL (POTENTIAL).
 FT DOMAIN 4359 4386 COILED COIL (POTENTIAL).
 FT NP_BIND 1865 1872 ATP (POTENTIAL).
 FT NP_BIND 2163 2170 ATP (POTENTIAL).
 FT NP_BIND 2537 2544 ATP (POTENTIAL).
 FT NP_BIND 2880 2887 ATP (POTENTIAL).
 SQ SEQUENCE 4568 AA; 521568 MW; 028B52684F3B1676 CRC64;

Query Match 4.6%; Score 102.5; DB 1; Length 4568;
 Best Local Similarity 17.2%; Pred. No. 2.2e+02;
 Matches 76; Conservative 68; Mismatches 149; Indels 149; Gaps 17;

QY 21 EPIYDAYVTALSSRHAVIT--PKGTSLKESRIKISDPQAN-----GWE-----YHI 67
 Db 2291 DELYKWFQVQWRNHLVFTNPPSGGLRER--ASTSPALFNRCVLMFGDMSNAALYQV 3048
 QY 68 SSNTEKREYVS-----RNPFLKATIFIVAVIQPTEAFDEIITISPGHSGSDT 120
 Db 3049 GSELTRTMDLDRDTYEGSVRLPSC-----LVPSQPTV----- 3082
 QY 121 ETKTSNGEKFLYSHRAITEVEKTLGSSAGLVSVATSLSHFIPNVISTNKDILHNV 180
 Db 3083 -----RDVAVNT 3089
 QY 181 AQIAHCAQK-----KIGSGFVATAIYGLIVRRFOPALINDVFOYLSDEPKFTEL 234
 Db 3090 LCLVHKTVQKFNEMETKGGHVMACTPRHFLDFIKQFV-SLFHKRSDLBEKIHNLIGL 3148
 QY 235 KKLIESMWEHRECTLPYGIKLLMGDVKGSSFTPKVSRVLQKKKEPRESSVYVDQIN 294
 Db 3149 NKISSTEBQVE---LQSKIKLSNLOEKKEANLKLKEMLDQQAEEKKRSEQLQ 3204
 QY 295 SANQFMKELAREMEKRYDDETYIKELDSVEPLTVAIKIRKGLALTQKSEVPI-- 352
 Db 3205 K---ELAEQLKQAKK-----KITVENDLQVE--AVAEQAQVAGIKKSQLYEVKSM 3233
 QY 353 -PDVQQLDRCCGELPGCVGVVPPAGAGYDAIYVLENOVGNFK--QKTLNPDYFNH 408
 Db 3254 SSPPVTVKLT-----LEAICILGENVGTWKAIARQVWMDQDFWTR 3294
 QY 409 VYVVDLEBQCBGVLEKPEPDI 430
 Db 3295 ILQFTELTLPPELIKQW-EKTI 3315

```

RESULT 32
ID YN91 ANASP STANDARD; PRT; 378 AA.
AC P46080; O53486;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein al13391.
GN AL13391.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasaomoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
RN [2]
RP SEQUENCE OF 1-122 FROM N.A.
RX MEDLINE=95189729; PubMed=7883709;
RA Bauer C.C., Buikema W.J., Black K., Haselkorn R.;
RT "A short-filament mutant of Anabaena sp. strain PCC 7120 that
RT fragments in nitrogen-deficient medium."
RL J. Bacteriol. 177:1520-1526(1995).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB003589; BAB74090.1; -
DR EMBL; U09240; AAA18028.1; -
DR EMBL; S76266; AAB33742.2; -
DR PIR; AH2104;
DR InterPro; IPR003724; CoBd CoBd Bturf.
DR Pfam; PF02572; CoBd CoBd Bturf. 2.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 82 83 GF -> DS (IN REF. 2).
FT CONFLICT 117 122 VAKGAI -> ASKARS (IN REF. 2).
SQ SEQUENCE 378 AA; 41340 MW; 43A96409A3EF31F9 CRC64;

Query Match 4.6%; Score 101.5; DB 1; Length 378;
Best Local Similarity 22.5%; Pred. No. 7.6; Indels 93; Gaps 14;
Matches 64; Conservative 36; Mismatches 92; GSSGYTBD 248

QY 153 LVSVAATSLSHPIPVISTNKDILHNAQIAHCVAKKIGSGFDV-----ATAIYGLIV 207
DB 86 LIDQVATGAEFFGPREIT-----PDRSEKGRMDVAKAISGLSVVVD 131
QY 208 YRRFOAL-----INDVFOVLESDEPKF-----PELKKLIESNWEK-----HEEC 249
DB 132 LDEINVLVDLGLLSVDEVGTLSKSKQELIATGAGAPQKLDLIDHSEKMPILHHPA 191
QY 250 T--LPGIKLMDGVKGGSETP-----KLVS-----VLQMKKEKPESSVVVD 291
DB 192 TELMTGIEIITYGAGKSTSLGKALQAIIGRGINPGSTRVLIQWMLK-----GSSGYTBD 248
QY 292 QUNSANLQPMKEL-----REMKERYSDPEYIKELDHSVEPLTVAIRN----- 335
DB 249 AAIISALRQSYPEVVDHRCGRDAIWMRNSQKELD-----YV--BAEKGWELAKVALISGLY 302
QY 336 ---IRKGLQALTKQSEVPIEDVQTOI-----LDRQGEIP 367
DB 303 KTIILDELNPVDLRLPVEPIVQALLRKPRDTEIIITGRCOMOP 347

```

```

RESULT 33
ID GAL1 FUSNN STANDARD; PRT; 389 AA.
AC Q8RHD0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Galactokinase (EC 2.7.1.6) (Galactose kinase).
GN GALK OR FN2107.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykides A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyriades N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
CC -1- CATALYTIC ACTIVITY: ATP + D-galactose = ADP + D-galactose 1-
CC phosphate.
CC -1- PATHWAY: Galactose metabolism; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE GHMP KINASE FAMILY. GALK SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE010514; AAL94191.1; -
DR HAMAP; MF_00246; -; 1.
DR InterPro; IPR000705; Galactokinase.
DR InterPro; IPR006204; GHMP Kinase.
DR InterPro; IPR006203; GHMPKase ATP.
DR Pfam; PF00288; GHMP_Kinases; 1.
DR TIGRPFAMs; TIGR00131; gal_kin; 1.
DR PROSITE; PS00106; GALACTOKINASE; 1.
DR PROSITE; PS00627; GHMP KINASES ATP; 1.
KW Transferase; Kinase; Galactose metabolism; ATP-binding;
KW Complete proteome.
FT NP_BIND 122 132 ATP (POTENTIAL).
SQ SEQUENCE 389 AA; 43753 MW; 784257843B31D98E CRC64;

Query Match 4.5%; Score 100.5; DB 1; Length 389;
Best Local Similarity 20.1%; Pred. No. 9.3; Indels 169; Gaps 24;
Matches 94; Conservative 61; Mismatches 143; GSSGYTBD 248

QY 5 FSAPKAFIA-----GGYLVLEPIYAYVATLSSRMHVAITPRTGSLKESRIKSSPQ 57
DB 22 FSPGRVNLIGEHRTDYGGFV-----FPCALDFGYTAAVVKR-----EDKTFPMYSKN 69
QY 58 FANGEMEWHISSNTEKPREVQSRINPFLATIFLYATIQPREADLRIIYSDGHSQ 117
DB 70 FEN-----LGIIE-----FNLNLLY-----DK 87
QY 118 EDTETKSSNGEKTFLYSRATTE-----VEKGLSSAGLGVVATSLSHPIP 167
DB 88 KDDMANYPKGVKTIIDRNKYKIDSGFDVLFQININGAGLSSASIEVLTAVIILDLDFLT 147
QY 168 NVISTNKDILHNV--AQIAH-----CYAOKKISG-----FDVATAIYGLIYVR 209

```

Db 148 DV-----DIIENVKNCQVAENKFIGVNSGIMDQAVGKGDKNAILLDCNTLKY----- 196
Qy 210 RFQPALINDVFOVLESDEPEKFTPELKLIESNMEEKHERCTLPYGIKILMGDVKGSSETP 269
Db 197 EYVPEKLVNMSIVINANTMK-----RGLADSKYNERRTSC----- 231
Qy 270 KLVSVLQWKKKEPESSVYVDQLNSANIQFMKELR-----EKREKYSDEPETYIKELDHS 325
Db 232 -----BEAVVVLNN-NEVNIKYGLVTEFEFKVHYITD-EBQLKRATHA 275
Qy 326 V-----EPVVAIKNIKRG-----LQALTQKSEVPIEDVQV-----QLDRCEIGECVCG 372
Db 276 VYENERRAKTAVLEFKDDIAEFKGLMNSKSHTSLRDYEVTGLDLSLVEAAMEEKGYVGS 335
Qy 373 VVPGAGGYDAIVLLEN-----QVGS-NFKOKT-LENPDYFHNV 409
Db 336 RMTGA-GRGGCTVSIENDYVDFIKNGVKYKENTGLEASFTYANI 381

RESULT 34
FTSY_AQUAE
ID FTSY_AQUAE STANDARD; PRT; 461 AA.
AC 067066;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell division protein ftsy homolog.
OS FTSY OR AQ 920.
GN Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
Graham D.E., Overbeek R., Shead M.A., Keller M., Aujay M., Huber R.,
Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus".
RL Nature 392:353-358(1998).
CC -1- FUNCTION: FUNCTIONAL HOMOLOGY OF SRP RECEPTOR. PROBABLY INVOLVED IN
THE RECEPTION AND INSERTION OF A SUBSET OF PROTEINS AT THE
CYTOPLASMIC MEMBRANE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; AE000714; AAC07030.1; -
DR PIR; G70379; G70379.
DR HSSP; P10121; 1FYS.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR004390; ABC_transp_Ftsy.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00448; SRP54_1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF00515; TPR; 1.
DR ProDom; PD000819; SRP54_1.
DR SMART; SM00382; AAA; 1.
DR TIGRfams; TIGR00064; ftsy; 1.
DR PROSITE; PS00300; SRP54_1.
KW Signal recognition particle; GTP-binding; RNA-binding; Membrane;
KW Cell division; Complete proteome.
FT NP_BIND 267 274 GTP (BY SIMILARITY).
FT NP_BIND 349 353 GTP (BY SIMILARITY).

FT NP_BIND 413 416 GTP (BY SIMILARITY).
SQ SEQUENCE 461 AA; 51930 MW; 9D66F143849A662F CRC64;
Query Match 4.5%; Score 100.5; DB 1; Length 461;
Best Local Similarity 24.5%; Pred. No. 12;
Matches 65; Conservative 43; Mismatches 88; Indels 69; Gaps 14;
Qy 177 LHNVAQIHCYCAOKKIGSGFVATAIYGLIYRRFPQPALINDVFOVLESDEPEKFTPELKK 236
Db 108 LYNVAKIYH--QLEKPDALLEYAQAELVYERK-----KELENFITQLKK 152
Qy 237 ---LIESNMEEKHERCTLPYGIKILMGDVKGSSETPKLVSVLQWKKKEPESSVYVDQL 293
Db 153 ELGLI-----EEKKE-----SILDKLRRLQKTEKAEVFGVLFGRKVD----- 192
Qy 294 NSANLQFMKELRREKYSDEPETYIKELDHSVERVVAIKNIKRGALQLOKSEVPIEP 353
Db 193 -----EPEELEEMLVKADVGVKTAV--ELTEKLRREAIR-KNIEG-----EKIKELKK 240
Qy 354 DVQOTQLDRQCO---EIPGCVGV-----VPGAGGYDAIVLLE-NQVG-----NF 395
Db 241 ELK-ELKNCQGLKIPKVGAVLLFVGVSNGTTTIGKLAHQKQKGVLLVAGDTF 299
Qy 396 KQKTELENDYFHNVYVDLEBQTEG 420
Db 300 RAAIEGLEVWAKGADVIVKKEG 324

RESULT 35
YQ91_CAEEL
ID YQ91_CAEEL STANDARD; PRT; 782 AA.
AC 010030;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical 90.1 kDa protein C27D6.1 in chromosome II.
GN C27D6.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Ding H.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; U23179; AAC46721.1; -
DR PIR; T15654; T15654.
DR WormPep; C27D6.1; CE01811.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 669 672 POLY-ALA.
FT DOMAIN 223 331 COILED COIL (POTENTIAL).
FT DOMAIN 348 398 COILED COIL (POTENTIAL).
FT DOMAIN 428 601 COILED COIL (POTENTIAL).
FT DOMAIN 699 743 COILED COIL (POTENTIAL).
SQ SEQUENCE 782 AA; 90059 MW; BB7CB842CD7373BE CRC64;
Query Match 4.5%; Score 100.5; DB 1; Length 782;
Best Local Similarity 21.1%; Pred. No. 25;
Matches 92; Conservative 61; Mismatches 158; Indels 125; Gaps 21;
Qy 22 PIYDAYVTALSSRMHAAVTPKGTSLKESRIKISSPQ--PA-NGEWEYHIS-----SNT 71
Db 38 PISAPFRTASDIGHHST-----KMYFKILSPNNVAFALMGDNEKTSFKLFTIPEPT 90

QY 72 EKPR-EVOS-----RINPFL---FATIFVLAVIOPTFAFDLEIITY-----SDPGY 114
 DB 91 STPNKGIQGLITITSYKINNVLFGMEKEIF-----IEKANI FSLHAIHFPQSTNRNPAS 145
 QY 115 HSQEDTETRTSSNGEKTFYHSAITEV--EKTGLSSAGLVSVATSLSHETPNVIST 172
 DB 146 ESHHQDLSSMALVSTLNRGNKITTTLRPRTPYKRAADGMSHEDLRS---PDVSAT 202
 QY 173 NMDIL---INVQIACVAGQKKGSGPDVATAYGLIVRRPPALINDVFOVLESDEPK 229
 DB 203 FTTVLVFKVFOVAATYTLRERSRQTKVESL---BRL-----K 243
 QY 230 PFTLKLIESNMEGHERCTLPYGIKLMDGVGSEFPKLYSRVLQWKEKPESSVY 289
 DB 244 ANEKARKEIEAEKAKKDRAT-----RNRKRLPELELELAETVQAK-EEWQYK 290
 QY 290 YQQLNSANTQFMKELREMEKDYDPEYIKELHDSVEPLTVAIKIRKGLQALTKQSEV 349
 DB 291 SQEMEIQNKQVLELNEVQK-----LEBIENSQKTFH-----QKVSTLNDDEY 336
 QY 350 PIEPDVOTQLDRCQRIPGCVGIVPGAGGYDAIYLVLENQVGNFKQKTLNPDYFNHV 409
 DB 337 FQNPDEEQD-----GSLSQFNMDVLEQELVKYQEK----- 367
 QY 410 YWVDLEQTEGVLEEK 425
 DB 368 --ISLKQENE-ILKEK 360

RESULT 36
 DMD_HUMAN
 ID P1532; Q14169; Q14170; PRT; 3685 AA.
 AC P1532; Q14169; Q14170; Created
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last annotation update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Dystrophin.
 GN DMD.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88194521; PubMed=3282674;
 RA Koenig M., Monaco A.P., Kunkel L.M.;
 RT "The complete sequence of dystrophin predicts a rod-shaped
 RT cytoskeletal protein.";
 RL Cell 53:219-228 (1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89345106; PubMed=2668885;
 RA Rosenthal A., Speer A., Billowitz H., Cross G.S., Forrest S.N.,
 RA Davies K.E.;
 RT "Two human cDNA molecules coding for the Duchenne muscular dystrophy
 RT (DMD) locus are highly homologous";
 RL Nucleic Acids Res. 17:5391-5391 (1989).
 RN [3]
 RP SEQUENCE OF 1-497 FROM N.A.
 RX MEDLINE=87273512; PubMed=3607877;
 RA Koenig M., Hoffman E.P., Bertelson C.J., Monaco A.P., Feener C.,
 RA Kunkel L.M.;
 RT "Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and
 RT preliminary genomic organization of the DMD gene in normal and
 RT affected individuals.";
 RL Cell 50:509-517 (1987).
 RN [4]
 RP SEQUENCE OF 404-1137 FROM N.A.
 RX MEDLINE=88115152; PubMed=3428261;
 RA Cross G.S., Speer A., Rosenthal A., Forrest S.M., Smith T.J.,
 RA Edwards Y., Flint T., Hill D., Davies K.E.;
 RT "Deletions of fetal and adult muscle cDNA in Duchenne and Becker

RT muscular dystrophy patients.";
 RL EMBO J. 6:3277-3283 (1987).
 RN [5]
 RP SEQUENCE OF 665-722; 2098-2204 AND 2305-2366 FROM N.A.
 RX MEDLINE=89083552; PubMed=3205741;
 RA Chamberlain J.S., Gibbs R.A., Ranier J.A., Nguyen P.N.,
 RA Caskey C.T.;
 RT "Deletion screening of the Duchenne muscular dystrophy locus via
 RT multiplex DNA amplification.";
 RL Nucleic Acids Res. 16:1141-1156 (1988).
 RN [6]
 RP SEQUENCE OF 2147-2204 FROM N.A.
 RX MEDLINE=89345155; PubMed=2569720;
 RA Blonden L.A.J., den Dunnen J.T., van Paassen H.M.B.,
 RA Wapenaar M.C., Gooschoten P.M., Ginjaar H.B., Bakker E.,
 RA Pearson P.L., van Ommen G.J.B.;
 RT "High resolution deletion breakpoint mapping in the DMD gene by whole
 RT cosmid hybridization.";
 RL Nucleic Acids Res. 17:5611-5621 (1989).
 RN [7]
 RP SEQUENCE OF 2305-2364 FROM N.A.
 RA Huth A., Will K., Speer A., Bauer D.;
 RL Submitted (MAR-1991) to the EMBL/Genbank/DBJ databases.
 RN [8]
 RP INTERACTION WITH SNTB1.
 RX MEDLINE=95146543; PubMed=7844150;
 RA Ahn A.H., Kunkel L.M.;
 RT "Syntrophin binds to an alternatively spliced exon of dystrophin.";
 RL J. Cell Biol. 128:363-371 (1995).
 RN [9]
 RP INTERACTION WITH SNTA1 AND SNTB2.
 RX MEDLINE=96162017; PubMed=8576247;
 RA Ahn A.H., Feener C.A., Gussoni E., Yoshida M., Ozawa E., Kunkel L.M.;
 RT "The three human syntrophin genes are expressed in diverse tissues,
 RT have distinct chromosomal locations, and each bind to dystrophin and
 RT its relatives.";
 RL J. Biol. Chem. 271:2724-2730 (1996).
 RN [10]
 RP INTERACTION WITH SNTG1 AND SNTG2.
 RX MEDLINE=20283612; PubMed=10247910;
 RA Piluso G., Mirabella M., Ricci E., Belisio A., Abbondanza C.,
 RA Servidi S., Puccia A.A., Tonali P., Puccia G.A., Nigro V.;
 RT "Gamma1- and gamma2-syntrophins, two novel dystrophin-binding proteins
 RT localized in neuronal cells.";
 RL J. Biol. Chem. 275:15851-15860 (2000).
 RN [11]
 RP REVIEW ON DMD POINT MUTATION VARIANTS.
 RX MEDLINE=95038763; PubMed=7951253;
 RA Roberts R.G., Gardner R.J., Bobrow M.;
 RT "Searching for the 1 in 2,400,000: a review of dystrophin gene point
 RT mutations.";
 RL Hum. Mutat. 4:1-11 (1994).
 RN [12]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=94320940; PubMed=8045556;
 RA Rimsland F., Reiss J.;
 RT "Microlesions and polymorphisms in the Duchenne/Becker muscular
 RT dystrophy gene.";
 RL Hum. Genet. 94:111-116 (1994).
 RN [13]
 RP VARIANT DMD ARG-54.
 RX MEDLINE=94004962; PubMed=8401582;
 RA Prior T.W., Papp A.C., Snyder P.J., Burghes A.H.M., Bartolo C.,
 RA Sedra M.S., Western L.M., Mendell J.R.;
 RT "A missense mutation in the dystrophin gene in a Duchenne muscular
 RT dystrophy patient.";
 RL Nat. Genet. 4:357-360 (1993).
 RN [14]
 RP VARIANTS DMD HIS-365; TRP-2191 AND ARG-2937.
 RX MEDLINE=95152525; PubMed=7849724;
 RA Nigro V., Nigro G., Esposito M.G., Comi L.I., Molinari A.M.,
 RA Puccia G.A., Pollano L.;
 RT "Novel small mutations along the DMD/BMD gene associated with

RT different phenotypes";
 RL Hum. Mol. Genet. 3:1907-1908 (1994).
 RN [15]
 RP ALTERNATIVE SPLICING (ISOFORMS 1 AND 2).
 RC TISSUE=Retina;
 RA White R.A.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [16]
 RP ALTERNATIVE SPLICING (ISOFORM 3).
 RC TISSUE=Brain;
 RX MEDLINE=89181947; PubMed=2648158;
 RA Feener C.A., Koenig M., Kunkel L.M.;
 RT "Alternative splicing of human dystrophin mRNA generates isoforms at the carboxy terminus";
 RL Nature 338:509-511 (1989).
 CC -1 FUNCTION: May play a role in anchoring the cytoskeleton to the plasma membrane.
 CC -1 SUBUNIT: Interacts with the syntrophins SNTA1, SNTB1, SNTB2, SNTG1 and SNTG2.
 CC -1 ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Comment=Additional isoforms seem to exist;
 CC Name=4; Synonyms=Dystrophin-4;
 CC IsoId=P11532-1; Sequence=Displayed;
 CC Name=1; Synonyms=Dystrophin-1;
 CC IsoId=P11532-2; Sequence=VSP_006806, VSP_006807;
 CC Name=2; Synonyms=Dystrophin-2;
 CC IsoId=P11532-3; Sequence=VSP_006807, VSP_006808;
 CC Name=3; Synonyms=Dystrophin-3;
 CC IsoId=P11532-4; Sequence=VSP_006809;
 CC -1 DISEAS: Defects in DMD are the cause of Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD). DMD is the most common form of muscular dystrophy; a sex-linked recessive disorder. It typically presents in boys aged 3 to 7 year as proximal muscle weakness causing waddling gait, toe-walking, lordosis, frequent falls, and difficulty in standing up and climbing up stairs. The pelvic girdle is affected first, then the shoulder girdle. Progression is steady and most patients are confined to a wheelchair by age of 10 or 12. Flexion contractures and scoliosis ultimately occur. About 50% of patients have a lower IQ than their genetic expectations would suggest. There is no treatment. BMD resembles DMD in hereditary and clinical features but is later in onset and more benign.
 CC -1 DISEAS: Defects in DMD are a cause of X-linked dilated cardiomyopathy (XLCM).
 CC -1 MISCELLANEOUS: The DMD gene is the largest known gene in humans. It is 2.4 million base-pairs in size and comprises 79 exons.
 CC -1 SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN, ABP-120, ABP-180, OR BETA-FODRIN).
 CC -1 SIMILARITY: Contains 2 calponin-homology (CH) domains.
 CC -1 SIMILARITY: Contains 22 spectrin repeats.
 CC -1 SIMILARITY: Contains 1 WW domain.
 CC -1 SIMILARITY: Contains 1 ZZ-type zinc finger.
 CC -1 DATABASE: NAME=DMD; NOTE=Dystrophin Mutation Database; WWW="http://www.dmd.nl/database.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/ or send an email to license@isb.ch).
 CC -----
 CC EMBL, X06179; CAA29545.1; ALT_SEQ.
 CC EMBL, X06178; CAA29544.1; -
 CC EMBL, X14298; CAA32479.1; ALT_SEQ.
 CC EMBL, X15495; CAA33518.1; -
 CC EMBL, X54820; CAA38589.1; -
 CC EMBL, X13045; CAA31451.1; -
 CC EMBL, X13046; CAA31452.1; -
 CC EMBL, X13047; CAA31453.1; -

DR EMBL, X13048; CAA31454.1; -
 DR EMBL, U27203; AAA86115.1; -
 DR EMBL, U27203; AAA86116.1; -
 DR EMBL, X15148; CAA33245.1; -
 DR PDB; 1DXK; 16-MAY-00.
 DR PDB; 1EG3; 23-AUG-00.
 DR PDB; 1EG4; 23-AUG-00.
 DR Genew; HGNC:2928; DMD.
 DR MIM; 300377; -
 DR MIM; 310200; -
 DR MIM; 300376; -
 DR MIM; 302045; -
 DR GO; GO:0005856; C:cytoskeleton; TAS.

Query Match 4.5%; Score 100.5; DB 1; Length 3685;
 Best Local Similarity 20.9%; Pred. No. 2.3e+02;
 Matches 102; Conservative 78; Mismatches 167; Indels 141; Gaps 27;

QY 14 AGYLVLEPIYDAYVTALSSRM---HAVITPKGTSLSKSRKIS---POPAN 60
 DB 849 AENNLIKQPTTPSEPTAIKQKICKOVNLSGLQPOIEKLIKQSLMKKGQPMFLD 908
 QY 61 GMEYHISNTEKPREVQSR-----INPF-LEATITIVLAYIOTPAFLLEITTY 109
 DB 909 ADFAFTNHFQVPSDYAREKEKLTITFDLPKRYQETMSAIRTWQOSET-KLSIPL 967
 QY 110 SDPGYHSGED-----TEKTSNGEKPELYH-SRAITVEKTVGLSSA-----G 152
 DB 968 SVTDEIWEQRLGLQALQSLQEQSGLYLSTTVKEMSKAPSEISRKYQSEFEIEG 1027
 QY 153 LVSVAITSLSHF-----IPNVISTNDILHNV-----AQ 182
 DB 1028 RKKKLSQSLVHCQKLEBQNMKLRKIQNHITLKKMALEVDVFLKEWPAIDSEILKKQ 1087
 QY 183 IAHCAQKKISGSDVATATIGLYRRFQAL-INDVFOYL-ESDPE--KEPTLKK 235
 DB 1088 LKQC--RLVVS--DIQT-----IOPSLSVNEGKIKINEAPEPASPRLTEILK 1132
 QY 236 KLISNMEKKEKRCPLPFGIKLMDGVKGSB---TPKLSRYLOWKKEPDE----- 285
 DB 1133 EL-NTQND--HMCQVYARKEAL--KGLSEKTVSLQKLSHMEHWTOABEETLEDF 1185
 QY 286 SSVVYDQLSANLQFMKELREMERKSDPEPTIYKELDHSEPL-----TVAIKNIRKGL 340
 DB 1186 EYKTPDELQKVEEMKKAKEAQAQK-----EAKYKLITESTNSVIAQAPPAQELKKEL 1240
 QY 341 QALTQKSEVPLEPDVQGLDRG---QETPGCVGVPPAGGYDAIAVLLENQVN--- 394
 DB 1241 ETLTNTYQW-----LCTRINKCKTLEEVNAC-----WHBILSYLERANKWINEVE 1286
 QY 395 FKQKTLRN 402
 DB 1287 FKLTETEN 1294

RESULT 37

GIDA_MYCGE STANDARD; PRT; 612 AA.
 AC P47619;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glucose inhibited division protein A.
 GN GIDA OR MG379.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2097;
 RN [1]
 RP SEQUENCE FROM N.A. / G-37;
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=96026346; PubMed=7569993;
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,

RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
 RA Nguyen D.T., Uterback T.R., Sauder D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RT "The minimal gene complement of Mycoplasma genitalium";
 RL Science 270:397-403(1995).
 (2)
 RN SEQUENCE OF 301-394 FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=94075230; PubMed=8253680;
 RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III;
 RT "A survey of the Mycoplasma genitalium genome by using random
 sequencing";
 RL J. Bacteriol. 175:7918-7930(1993).
 CC -1- FUNCTION: NOT KNOWN.
 CC -1- SIMILARITY: Belongs to the gida family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U03719; AAC71606.1; -;
 CC EMBL; U01812; AAD12347.1; -;
 CC PIR; I64241; I64241.
 CC TIGR; MG379; -;
 CC HAMAP; MF_00129; -; 1.
 CC InterPro; IPR002218; GIDA.
 CC InterPro; IPR004416; Gida sub.
 CC InterPro; IPR000205; NAD_binding.
 CC Pfam; PF01134; GIDA; 1.
 CC ProDom; PD003738; GIDA; 1.
 CC TIGRPFAM; TIGR00136; gida; 1.
 CC PROSITE; PS01280; GIDA_1; 1.
 CC PROSITE; PS01281; GIDA_2; 1.
 CC Complete proteome.
 CC KW Complete proteome.
 CC SEQUENCE 612 AA; 68995 MW; 67A3F64C985B5912 CRC64;
 SO
 Query Match 4.5%; Score 100; DB 1; Length 612;
 Best Local Similarity 23.1%; Pred. No. 19;
 Matches 88; Conservative 59; Mismatches 158; Indels 76; Gaps 19;
 QY 4 AASAPKATLAGYVLEPIYAYVATLSRMAVITP--KGTSLKSKIKTSSPPFANG 61
 DB 222 AESTNNKFLP-----LEKQIPCYLVHTNQKHDLIKVLKKSAMFNGSISAGPLVC-P 275
 QY 62 EMEYHSSNTEKREYQSRINPFLKATIFVLAYIQTEAPDEI---IYSDPGYHQS 118
 DB 276 SIEDKVPKFSQKRRH-QIFVEPESLSLDITLNGL--STSFPEIQKEIIQLPFGQNH 332
 QY 119 -----DTEKTSNGEKTFLYHSRAITEV---EKTGLGSGAGLV 154
 DB 333 IKKFGVAIEYDAFLSNQKLPLETLEKLIEN-----LTFAGQINGTSGYEAAAGGLMAGIN 387
 QY 155 SVAVTSLSHFIPNVSTNKDILAHVYQAQKAGKISGFPVAT--AIYGLIYVR-RF 211
 DB 388 A--ALTLKK-PPIIIOQNEAV---IGVMINDLVYITISDPYRLTISRAREYRLMLNDNV 441
 QY 212 QPALINDVQVLESDEKPFTELKLIESNWEKHERCTLPYGIKLMDVYGSGSPPL 271
 DB 442 QERLIKSSSELGLTDKTYELFLK-----EKKQELISFLAKNTQGVVAKLKTNNK 494
 QY 272 VSRVLQWKKEKRESSVVDQNLANSANIQFKELREMRKXKSDPETYIELDHSVEPLTV 331
 DB 495 TNGSL-----YDFNKRSEINIDKLIKIDPKKYQDSEI-LKQIEIEIKYEGY 540
 QY 332 AIKNIR--KGIQALTQKSEVP 350
 DB 541 IKNEKYFKGLDKLS-KIRIP 560

RESULT 38
 ID Y011 YEAST
 AC P47108; STANDARD; PRT; 1174 AA.
 DT 01-FEB-1996 (Rel. 33; Created)
 DT 01-FEB-1996 (Rel. 33; Last sequence update)
 DT 15-DEC-1998 (Rel. 37; Last annotation update)
 DE Hypothetical 135.1 kDa protein in GEF1-NUP85 intergenic region.
 GN Y0R041C OR Y1622.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8288C;
 RX MEDLINE=95397595; PubMed=7668047;
 RA Huang M.-E., Chast J.-C., Galibert F.;
 RT "Analysis of a 42.5 kb DNA sequence of chromosome X reveals three
 RT rRNA genes and 14 new open reading frames including a gene most
 RT probably belonging to the family of ubiquitin-protein ligases";
 RL Yeast 11:775-781(1995).
 CC -1- SIMILARITY: SOME, TO S.POMBE SPAC2G11.02.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L36344; AAA88743.1; -;
 CC EMBL; Z49541; CAA89568.1; -;
 CC EMBL; Z49542; CAA89570.1; -;
 CC PIR; S57060; S57060.
 CC SGD; S0003802; Y0R041C.
 CC KW Hypothetical protein.
 CC SEQUENCE 1174 AA; 135117 MW; 7647D48DAS57DAD4 CRC64;
 SO
 Query Match 4.5%; Score 100; DB 1; Length 1174;
 Best Local Similarity 22.0%; Pred. No. 48;
 Matches 85; Conservative 63; Mismatches 123; Indels 116; Gaps 21;
 QY 62 EME--YHSSNTEKREYQSRINP--LEATIFVLAYIQTEAPDEIITISDGYHSQ 117
 DB 480 QMDLYTHI-----MEYVDIVPAEIEKIDYVLSNIFPITSADVELEFPY---CEFL 529
 QY 118 ETEKTSNGEKTFLYHSRAITEVEKTLGSSAGLVAVATSLSHFIPNVIST----- 172
 DB 530 KEVTFDLSDAKKKRRFELIDERSNL--SYSVKRFVLVNNNTREQISSLIISL 587
 QY 173 -----NKDILH--NVAQIANCYAQ--KKISGFPDVAATAYGI-----VYERF 211
 DB 588 LINSTNLISLKNDDIFPEVTIT-----YALINKLASSYHGFALNALIQPIQINQV 642
 QY 212 QPALIND-----VFOVLESDEKPFTELKLIESNWEKHERCTLPYGIK 256
 DB 643 RVALINNTLCESPCLDSATRECLTLALS-----PT-FKSNIEITNP---YELCE-----K 688
 QY 257 LIMGVKGSGSETPKLVSVNLQWKKEKRESSVVDQNLANSANIQFKELREMRKXKSDP 316
 DB 689 TMSPEMAISETGD-----EKKEIDKISIFKAVTNNHSAQKE--PVSSEKLESGY 738
 QY 317 TYIKE--LDHSVEPLVAIRKIRKQALTQKSEVPLEPVGQTQLDRCOEIPCGVGV 373
 DB 739 DIVQSMVLSNGDSKLLIAGFTIAFL-----KPD-----NKHRIQDMA--- 778
 QY 374 VEGAGYDAIYAVLVLENQGNPKQTL 400
 DB 779 -----ISYAVKTIENYSENFESETI 798


```

RESULT 39
ABRA PLAF
ID ABRA PLAF STANDARD: PRT; 321 AA.
AC P33746;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE 101 kDa malaria antigen (P101) (Acidic basic repeat antigen)
DE (Fragment).
GN ABRA.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87038741; PubMed=354513;
RA Stahl H.-D., Bianco A.E., Crewther P.B., Anders R.F., Kyte A.P.,
RA Coppe R.L., Mitchell G.F., Kemp D.J., Brown G.V.;
RT "Sorting large numbers of clones expressing Plasmodium falciparum
RT antigens in Escherichia coli by differential antibody screening.";
RT Mol. Biol. Med. 3:351-368 (1986).
CC -1- SUBCELLULAR LOCATION: AT THE MEROZOITE SURFACE AND WITHIN THE
CC PARASITOPHOROUS VACUOLE.
CC -1- PTM: NOT GLYCOSYLATED (PROBABLE).
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL; M15514; AAA29463.1; -
DR Antigen; Malaria; Repeat.
FT NON TER 1
FT DOMAIN 275 >321 TANDEM REPEATS OF K-E(2,3).
FT NON TER 321
FT SEQUENCE 321 AA; 37805 MW; 0BBB4A0681F0714B CRC64;
SQ
Query Match 4.4%; Score 99; DB 1; Length 321;
Best Local Similarity 21.5%; Pred. No. 9;
Matches 58; Conservative 37; Mismatches 75; Indels 100; Gaps 11;
QY 228 EKEPPELKLIESNMEKHEKRTLPYGTILMGVKGSGEPTKLVSRVLQMKK----- 281
DB 20 EYVEELKKGI-----LDMGKILFSSKVK-SLKKLKNKILFPKKKEDNQAVD 65
QY 282 -----KPESSVYVQDLSNANLQ-----FMKELE----- 306
DB 66 TKSMEPKYKQAPALRGVEPTEDSNINMSINNVMDEIDFEKELLENNTPRVNVPPTQSK 125
QY 307 -----MKKYDSDEPTYIK-----LDHSEVPLTVAIK-----NIRKG 339
DB 126 KKNKKEIVSGMDENHPENYFKEEYDYDENDMEVKKIGVTLKKEPPLKKNVSET 185
QY 340 LQAL-----TKSEVPFIEPDVOTQLDRQCEIFGCVGVVPAGGYDAIVLVLENQGN 394
DB 186 IKLHGNKKDKKHIAINNDIQIK-----QELQAIYNELMNTYNGKNQIQI----- 233
QY 395 FKQKTLNPDVYFHYVWVLEQTEGVLEE 424
DB 234 FQGNLTLEN-DVLTNQTETEEEMKQVATIKQ 262

```

```

RESULT 40
M3K9 HUMAN STANDARD: PRT; 394 AA.
AC P80152;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

```

```

DE Mitogen-activated protein kinase kinase kinase 9 (EC 2.7.1.1-) (Mixed
DE lineage kinase 1) (Fragment).
GN MAP3K9 OR MLK1 OR PRKEL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Colon epithelium;
RA MEDLINE=93238756; PubMed=8477742;
RA Dorow D.S., Devereux L., Dietzsch E., de Kretser T.;
RT "Identification of a new family of human epithelial protein kinases
RT containing two leucine/isoleucine-zipper domains.";
RT Eur. J. Biochem. 213:701-710 (1993).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN EPITHELIAL TUMOR CELL LINES OF
CC COLONIC, BREAST AND OESOPHAGEAL ORIGIN.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MAP KINASE KINASE KINASE SUBFAMILY.
CC PIR; S32467; J00229.
DR HSSP; P12931; 1FMK.
DR Gensw; HGNC:6661; MAP3K9.
DR MIM; 600136; -
DR GO; GO:0005524; F:ATP binding activity; NAS.
DR GO; GO:0004708; F:MAP kinase kinase activity; NAS.
DR GO; GO:0006468; P:protein amino acid phosphorylation; NAS.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser_Thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; Pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Prot kinase; 1.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferrase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
KW ATP-binding.
FT NON TER 1
FT DOMAIN 3 271 PROTEIN KINASE.
FT NP_BIND 9 17 ATP (BY SIMILARITY).
FT BINDING 30 30 ATP (BY SIMILARITY).
FT ACT_SITE 127 127 BY SIMILARITY.
FT DOMAIN 289 310 LEUCINE-ZIPPER 1 (BY SIMILARITY).
FT DOMAIN 324 345 LEUCINE-ZIPPER 2 (BY SIMILARITY).
FT DOMAIN 354 368 ARG/LYS-RICH (BASIC).
FT SEQUENCE 394 AA; 44975 MW; DBE40B7D31047F08 CRC64;
SQ
Query Match 4.4%; Score 99; DB 1; Length 394;
Best Local Similarity 20.1%; Pred. No. 12;
Matches 87; Conservative 60; Mismatches 169; Indels 116; Gaps 17;
QY 10 KAPLAGGYLV-----EPIYDAYVTALSSRMHA-----VITPKGSLKESRIKIS 54
DB 19 RAFWIGDEVAVKARHPDEDDISQTIENVRQBAKLFMLKHPNIIALRGVCLKEPNLCLV 78
QY 55 SPQFANGMEWHISNTEKPREVQ-----SRINPL--EATFTV-----LAYIQ 97
DB 79 M-EFARGGPLNLRVLSGKRIPDILLVMAVOIARKNVLAHDAIVPIIHRDKSSNIIILQ 137
QY 98 PTEAFDL--EIIISDGYHSQEDTETKTSNGEKTFLYSHRAITVEKTLGSSAGLVS 155
DB 138 KVENGDLSNKLTKITDGLAREMERTKMSAG--TYAMAPVPIYRASMFKSGSDWMSYG 195
QY 156 VVATSLLSHPIPNVISNKKDILHNVAQIANCYAAKIGSGFDVATAYGLIVRRFOPAL 215
DB 196 VILWELLTGEVP-----FRGIDGLRVAVGVAMNKLALPI- 229
QY 216 INDVFOVLESDEPKFPPELKLIESNWE-EKHEKTLDPYG-KILMGVKGSGSTPLKVS 273
DB 230 -----PSTCPFPFALMEDCWNPPHSPSPFTNILDQLTTIEESGFFPMPDQSF 278
QY 274 RVLQ--WKKEKPESSVYVQDLSNANLQFMKELEMEKREKYSDEPTYIKELDHSVEPLTV 331

```

Db	279	HCLQDNWKHEIOE	----	MFQULAKE	----	KELR	-----	TWEEELTRAAALQOKN	319
Qy	332	AIKNIRKGIQALTQKSEVPIEPDVOTOLDRCOEIPGCVGVPAGGYDAIAVLVLNO	391						
Db	320	QEBLARRRQOELAREIDILEREINIHHQLCOEKPR	----	YKCR	360				
Qy	392	VGNFKOKTLENP	403						
Db	361	KGKFRKSRLAQP	372						

Search completed: August 8, 2003, 09:37:56
Job time : 30 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comogen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2003, 09:35:51 ; Search time 29 Seconds
(without alignments)
630.286 Million cell updates/sec

Title: US-10-069-062-7

Perfect score: 2230
Sequence: 1 MSKAPGAKFLAGYLV.....DLBQTEGLERKPDYIGL 432

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/6C_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	181.5	8.1	362	US-09-134-001C-5355	Sequence 5355, Ap
2	177.5	8.0	365	US-09-107-532A-6590	Sequence 6590, Ap
3	111.5	5.0	740	US-08-257-073-5	Sequence 5, Appl
4	111	5.0	2662	US-09-595-684B-31	Sequence 31, Appl
5	108	4.8	797	US-09-198-452A-636	Sequence 636, App
6	105	4.7	816	US-09-107-532A-3862	Sequence 3862, Ap
7	99	4.4	394	US-09-345-473E-19	Sequence 19, Appl
8	97.5	4.4	506	US-09-134-001C-4383	Sequence 4383, Ap
9	96.5	4.3	347	US-09-495-406-13	Sequence 13, Appl
10	96	4.3	3878	US-09-914-259-11	Sequence 11, Appl
11	95.5	4.3	342	US-09-107-532A-5664	Sequence 5664, Ap
12	95.5	4.3	520	US-09-030-995-3	Sequence 3, Appl
13	94.5	4.2	570	US-08-969-046-4	Sequence 4, Appl
14	94	4.2	646	US-09-107-532A-5713	Sequence 5713, Ap
15	94	4.2	3443	US-08-416-603-2	Sequence 2, Appl
16	93.5	4.2	1838	US-09-120-663-2	Sequence 17, Appl
17	92.5	4.1	1447	US-09-376-330-17	Sequence 6, Appl
18	92.5	4.1	2482	US-08-328-254-6	Sequence 1, Appl
19	92.5	4.1	3248	US-08-353-700-1	Sequence 1, Appl
20	92.5	4.1	3248	PCT-US95-16216-1	Sequence 1, Appl
21	92	4.1	1388	US-09-572-191-2	Sequence 2, Appl
22	92	4.1	1388	US-09-723-262-2	Sequence 2, Appl
23	92	4.1	1388	US-09-723-219-2	Sequence 2, Appl
24	91.5	4.1	280	US-09-328-352-5385	Sequence 5385, Ap
25	91.5	4.1	453	US-09-107-532A-6075	Sequence 6075, Ap
26	91.5	4.1	878	US-09-134-001C-4378	Sequence 4378, Ap
27	91	4.1	1156	US-09-198-452A-171	Sequence 171, App

28	90.5	4.1	803	1	US-08-062-368-2	Sequence 2, Appl
29	90.5	4.1	945	4	US-09-198-452A-1030	Sequence 1030, Ap
30	90	4.0	823	4	US-09-107-532A-6343	Sequence 6343, Ap
31	89.5	4.0	244	4	US-09-328-352-4409	Sequence 4409, Ap
32	89	4.0	517	4	US-09-257-825B-21	Sequence 21, Appl
33	88.5	4.0	304	4	US-09-134-001C-5153	Sequence 5153, Ap
34	88.5	4.0	1005	4	US-09-328-352-4877	Sequence 4877, Ap
35	88	3.9	437	4	US-09-328-352-5102	Sequence 5102, Ap
36	88	3.9	914	3	US-09-425-383-2	Sequence 2, Appl
37	88	3.9	1114	4	US-09-107-532A-5256	Sequence 5256, Ap
38	87.5	3.9	181	4	US-09-252-991A-21217	Sequence 21217, A
39	87.5	3.9	258	4	US-09-107-532A-3929	Sequence 3929, Ap
40	87.5	3.9	708	4	US-09-643-597-369	Sequence 369, App
41	87.5	3.9	976	3	US-09-104-324B-4	Sequence 4, Appl
42	87	3.9	396	6	5349058-2	Patent No. 5349058
43	87	3.9	641	3	US-08-961-083-160	Sequence 160, App
44	87	3.9	641	4	US-09-536-784-160	Sequence 160, App
45	87	3.9	1231	4	US-08-714-741-41	Sequence 41, Appl

ALIGNMENTS

RESULT 1
US-09-134-001C-5355
; Sequence 5355, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5355
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5355

Query Match	8.1%	Score 181.5;	DB 4;	Length 362;
Best Local Similarity	24.1%	Pred. No. 6.3e-09;		
Matches	99;	Conservative	52;	Mismatches 131; Indels 129; Gaps 18;
QY	7	APGKAFLLAGYLVLEPIYDAVYVTLASRMHAYI-----TPKGTSLKES	49	
DB	10	APGKLYIGEAIVATVPYKSLIIIAVRFVITAEASNVESGISHKTHYEVKDRND	69	
QY	50	RIKISSPOFANGEMWEYHISSTNTEKREVOSRINPELEATIFVLAYIOPTFADELIITY	109	
DB	70	RIEISDVQAQK-----LKVVAIAIEVEQYVNSCNML---	103	
QY	110	SDPGHSEDEPTETKSSNGEKTFLYHSRAITEVEKTEGSSAG--LVSVAASLSLHPIPN	168	
DB	104	-KHHHLTIDNLADNSG-----QRYGSSAAVAVSVV--KALNEFYG-	143	
QY	169	VISTKDIKHNVAOIAHCYAAOKKIGSGFDVATAY--GLIVRRFPQALIND--VPQVLSD	226	
DB	144	-LELSNLTYIKALAVYANKLOSLSGCG--DIASVYSGLAYSTFP-----HDVVKQOMER-	196	
QY	227	PEKPEPELKKLIESNWEKHE-ERCTLPYGIKLMDGVKSGSETPLVSRVLQWKKKEKBE	285	
DB	197	-----TSVNDVLEKRWPGIHEPLQAPENMEVLIGWTSPPASSPHLVSEVKKLK-----	245	
QY	286	SSVYVDQANSANLQFMKELREKREKYDDPEYIKELDHS---VEPLVAIK--NI-----	336	
DB	246	-----SDPSFYGDIFLDQSHACVESLIQAKRTNNIKGVQ	278	

```

Qy      337  -----RKGLQALTKSEVPIEDVDTQLDLRCOEIPGCVGVVPAGGYD 381D
          |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      279  KMIRINRRIQSMDNSEASYEIEITDKLKLCDVEKHGG--ASKTSGAGGGD 327T

```

RESULT 2

```

US-09-107-532A-6590
: Sequence 6590, Application US/09107532A
: Patent No. 6583275
: GENERAL INFORMATION:
: APPLICANT: Lynn A Doucette-Stamm and David Bush
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
: NUMBER OF SEQUENCES: 7310
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: GENOME THERAPEUTICS CORPORATION
: STREET: 100 Beaver Street
: CITY: Waltham
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02354
: COMPUTER READABLE FORM:
: MEDIUM TYPE: CD-ROM ISO9660
: COMPUTER: PC
: OPERATING SYSTEM: <Unknown>
: SOFTWARE: ASCII
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/107,532A
: FILING DATE: 30-Jun-1998
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/085,598
: FILING DATE: 14 May 1998
: APPLICATION NUMBER: 60/051571
: FILING DATE: July 2, 1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Arinello, Pamela Deneke
: REGISTRATION NUMBER: 40,489
: REFERENCE/DOCKET NUMBER: GTC-012
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (781) 893-5007
: TELEFAX: (781) 893-8277
: INFORMATION FOR SEQ ID NO: 6590:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 365 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: YES
: ORIGINAL SOURCE:
: ORGANISM: Enterococcus faecium
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: (B) LOCATION 1...365
: SEQUENCE DESCRIPTION: SEQ ID NO: 6590:
: US-09-107-532A-6590

```

[illegible]

RESULT 3

US-08-257-073-5
Sequence 5, Application US/08257073
Patent No. 576597
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: de Taisne, Charles
APPLICANT: Tine, John A.
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtiss, Morris & Safford, P.C.
STREET: 530 Fifth Avenue, 25th Floor
CITY: New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTMS
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 740 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-257-073-5

QY 126 SNGEFTLYHSRAITE---VEKTGIGSAGL-----VSVAVNSLSH-FIPNVISTN 173
Db 322 SUNDLINFLOQGEKEIGENILNITKMNIGDKNNLESFPLDQNMKNLNYEFILNLIKTS 381
QY 174 -----KDLHANAQIAH-CYAOKKIGSGPDVATAIYGLIVRRFPALINDVFOVLESBP 227
Db 382 VLNKLDLRLLYRAYVSYKRRKAOE-----KGLP-----EFTVTN----- 418
QY 228 EKFPTELKLIIESNBEKEHRCCTLPYGIKILMGDVKGSETPPLVSRVLOMKKE----- 281
Db 419 BEYVELEKKGIT-----LDMGITLFSKYK--SLKKLKNKKIPPKKEDNOAVD 464
QY 282 -----KPESSVYVDQNSANLQ--FMKELE----- 306
Db 465 TKSMEEPKVKAPALRGVFTEDSNIMNSINNVMDEIDFEKELIENNNTPVNVPPTQSK 524
QY 307 -----MREKXDSDEPTYKE-----LDHSVEPLTVAIK-----NIRKG 339
Db 525 KKNKNETVSGMDENDNHNEYFKEEYYDENDMEVKVKKIGVTLKKEPKNGVSEST 584
QY 340 LQAL-----TQSEVPIEPDVOTQLIDRCQEIIPGCGVGVPGAGYDAI AVLLENQVGN 394
Db 585 ICLIHGNKDKKHIAINNDIOLIK-----QELQALYNELMNTYTNKKNIOQL----- 632
QY 395 FKOKTLENDPYPHNVYWDLESEOTEGVLEE 424
Db 633 FOONILEN-DVLANQTEEMEKEQVEAITRO 661

RESULT 4

US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Berard, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Valseberg, Bugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; TITLE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; PRIOR FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: prt
; ORGANISM: Human
US-09-595-684B-31

Query Match 5.0%; Score 111; DB 4; Length 2662;
Best Local Similarity 20.5%; Pred. No. 0.96;

Matches 83; Conservative 66; Mismatches 163; Indels 92; Gaps 15;

QY 47 KESRIKISSPOFANGEMWHYHSSNTEKPREVOSRINPLEATIFIVLAIOTPEAFDEI 106
Db 852 KEAQFDSISLALKTLSKYTOELOKRETVQERINE-MEQLKEQLENNDSPLQTVREK 910
QY 107 IYSPRGHSDQEDTSTSSNGEKTFLYHSRAITEVEKTKGLSS----- 150
Db 911 TLITELQOTLEEVKTLVQ--EKDQLQLOESLQIERDQLSKSDHDTVMNIDTQEQLR 967
QY 151 ---AGLVSVATSLSHFIPNVISTNKDILHNAVAQIAHCYAOKKIG-----SGFDV 198
Db 968 NALLESKQETINTLKSXTSEVSNNLHNEBNTGETKDEFOQKXMGIDKKQLEAKNTOT 1027
QY 199 ATA-----YGLIVRRFPALINDVFOVLES---DEKPETELKLIIESNW 242

Db 1028 LTADVKNELIEQOKRIFSLIOEK-----NELQOMLESVIAEKEOLTKDIKENIEMTI 1080
QY 243 BEKEHRCCTLPYGIKILMGDVKGSETPPLVSRVLOMKK---EKPESSVYVDQNSANLQ 299
Db 1081 ENQEE-----LRL-----GDELKKQOEIVAOEKHAIKKEBELSRTCDRLAVEK 1127
QY 300 FMKELEHREKRYDS--DEPTYIKELDSVEPLTVAIKIRKGLQALTOKSEVPIEPDVOT 357
Db 1128 LKESQOQOEKQOQLNAVQEBSEWQKKIN---EIEMLKNEL-----KNKELTLB-HMET 1178
QY 358 QLLDRQEIIPGCGVGVPGAGYDAI AVLLENQVGNFKOKTLE 401
Db 1179 ERLBLAQKL-----NENYEBVKSITYERKVLKELQXSF 1212

RESULT 5

US-09-198-452A-636
; Sequence 636, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifflais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 636
; LENGTH: 797
; TYPE: prt
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 1...797
; OTHER INFORMATION: Xaa=unknown or other
US-09-198-452A-636

Query Match 4.8%; Score 108; DB 4; Length 797;
Best Local Similarity 19.7%; Pred. No. 0.26;

Matches 75; Conservative 69; Mismatches 133; Indels 104; Gaps 18;

QY 36 HAVITPKGTSKESRIKISSPOFANGEMWHYHSSNTEKPREVOSRINPLEATIFIVLAI 95
Db 226 YVIT--GISAQSPRIK-----QESLQALQKQD-----INAIVIDITNIMLSL 267
QY 96 IQTEAFPLEITIIYSDPGYHSDQEDTETSTSSNGEKTFLYHSRAITEVEKT--GLSSAG- 152
Db 268 GQPLHAYDASHVALDSLVEKLTSPESLTLINGETVLLPISGVVYVRDHSILGLGQWGA 327
QY 153 ---LVSVAATSL--SHFIPNVISTNKDILHNAVAQIAHCYAOKKIGSGPDVATAIYGLI 206
Db 328 KAPSFQETTTTVIKAAFLPEBALASQKLPIPESAYRFR-----GIDPQNVV----- 378
QY 207 VYRRFPALINDVFOVLESDEPKF-----FTBLKLIIESNMEKH 246
Db 379 -----PALQAIHNYHLEIFPEATISPIYSSGEICRELKEVALRKTLQRLIGKS----- 427
QY 247 ERTCLPYGIKILMGDVK--GSESTPK--LYSRVLOMKKEKPESSVYVD--QLNSANLQ 299
Db 428 ---FSIEILSQLOSLGFTTPOETSILVAKVPSYHNDINEEDLVEEICRTESWNIE 481
QY 300 FMKELEHREKRYDSDPT-----YIKELDSVEPLTVAIKIRKGLQALTKQS 347
Db 482 TQNPVSCCTPIYKLRKREAGFLANGLQOEFTTDDL--LDEIVTAL--TRK-----EKE 530
QY 348 EVPIEPDVOTQLDRQEIIPG 368
Db 531 EISLQSKHTITVL--RSSILPG 550

RESULT 6

US-09-107-532A-3862

Sequence 3862, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Denek
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3862:
SEQUENCE CHARACTERISTICS:
LENGTH: 816 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (8) LOCATION 1...816
SEQUENCE DESCRIPTION: SEQ ID NO: 3862:
US-09-107-532A-3862
Query Match 4.7%; Score 105; DB 4; Length 816;
Best Local Similarity 21.1%; Pred. No. 0.51;
Matches 76; Conservative 63; Mismatches 115; Indels 106; Gaps 19;
94 AYIQPEAPDL-----EIIYSDPG--YHSQEDTEKTSNGEKTFLYHSRAITEVEKTG 146
32 SFLSPKQYQVHNPFQLIAQNRGKIHAIQATSESSIRATQERSKKEITGRTEVAKR 91
147 LG-SSAGIVSVATISLH--FIP-NVISTNKDI-LHNAQIAHCAQKIGSGPVAARA 201
92 IDLTQALVSMIGDSQIDYRFIFGFKLIATDEEVNKNL-----KKSFFSGFO--EP 141
202 IYGLIYVRFQPLINDVQVLESDEKPEPTLKLIESNWEKHE-RCTLPYGIKLMG 260
142 VYGV-----NHLMD-FVLSLNEIRRTYKLEKMESTLARFRVRYRTPSDIAYLIE 194
261 DVKGSSETP-----KLVSRYLQWK-----KEKP-----ESSVY 289
195 HIYGEKGTPEEYEQPKKIKSETLVKRYDLLRSPRCLEKPRYLQMEHNEHSYVA 254
290 YQOLN-----SANTQFM-----KELDEM-R 308
255 YLTINIVGEMEPSSSELPYQOQFTFPIDTSMNVEIYVNNKALATVRNKKKELDLDN 314

309 EKYDSDPEYIKELD--HSEPLTVAIKIRKGLQALQKSEVPPEPDVQTLDRCOEI 366
315 HAYQSDNETNSVNLALDSVDELFTTLDQSKESMYKLSY--VVRVSASVDELKRCDEY 372
RESULT 7
US-09-345-473E-19
Sequence 19, Application US/09345473E
Patent No. 6558903
GENERAL INFORMATION:
APPLICANT: Hodge, Martin
TITLE OF INVENTION: No. 6558903el Kinases and Uses Thereof
FILE REFERENCE: 35800/183781
CURRENT APPLICATION NUMBER: US/09/345,473E
CURRENT FILING DATE: 1999-06-30
NUMBER OF SEQ ID NOS: 62
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 19
LENGTH: 394
TYPE: PRT
ORGANISM: Homo sapiens
US-09-345-473E-19
Query Match 4.4%; Score 99; DB 4; Length 394;
Best Local Similarity 20.1%; Pred. No. 0.58;
Matches 87; Conservative 60; Mismatches 169; Indels 116; Gaps 17;
10 KAPFLAGVYL-----EPYDAYVVALSRRMA-----VTPKQTSKESRIKS 54
19 RPFWIGDEAVAAAHDPDEDISQTIENVROEAKLFAMAKHPNITALKRVCLEKPEVLCIV 78
55 SQPANGEWEMHISNTEKPREVO-----SRINPFL--EATIRIV-----LAVIQ 97
79 M-EFARGGELNVLGSKRIPPDILVNNVAVQARGNNVYHDEALVPIIHRDLKSNILIQ 137
98 PTEAPDL--EIIYSDPGYHSQEDTEKTSNGEKTFLYHSRAITEVEKTGIGSSAGLVS 155
138 KYENGDLNKKIKITDIFGLAREMHTTWSAAG--TYAMMADEVIRASNFSGSVWSYSG 195
156 VVATISLHSPFNIVSTNKDILHNAQIAHCAQKIGSGPVAARAIIYRRFPQPL 215
196 VILMELTGEVP-----FRGIDGLVAAVANNKALPL- 229
216 INDVQVLESDEKPEPTLKLIESNWE-EKERCTLPYGI-KLMGDKGSGSETPKLV 273
230 -----PSTCEPPAKLMEDCQNPDPHSRSTFNILDQTLTIESGFFEMPKOSF 278
274 RVLQ--WKKKEPSESSVYDQLSANLQPMKGLREMRREKTDSDPEYIKELDHSVEPLTV 331
279 HCLQDNKHEIOE---MEDQLRAKE---KELR-----TWBEELTRALQOQKN 319
332 AIKIRKGLQALQKSEVPPEPDVQTLDRCOEIPGCVGVVPPAGGDAVAIVLENO 391
320 QEELLRREQELAEIEIDILEREINIITHQLCEKPR-----VKKR 360
392 VGNPKQKLENP 403
361 KKGFRKSRLAQF 372
RESULT 8
US-09-134-001C-4383
Sequence 4383, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08

```

? PRIOR APPLICATION NUMBER: US 60/055,779
? PRIOR FILING DATE: 1997-08-14
? NUMBER OF SEQ ID NOS: 5674
? SEQ ID NO 4383
? LENGTH: 506
? TYPE: PRY
? ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4353

```

Query Match	4.4%;	Score 97.5;	DB 4;	Length 506;
Best Local Similarity	25.4%;	Pred. No. 1.2;		
Matches 67;	Conservative 35;	Mismatches 99;	Indels 63;	Gaps 14;

Qy	43	GTSLKSRIRITSSPPCANEMEMEHJISNTEKREVOASINPFLKTIPIVLAIOPEAF	102
Db	7	GGRLMELVOKLSRQYIDGEM---VESSKNKTRDI---INPNQETIITV---AEGRKE-	56
Qy	103	DLEITIIYS-----DPSGHSQEDTETKTSNGEKTELPHYSRAITEVEKTLGSSAGLVSV	157
Db	57	DVRRALIAARRSPDEDMSELTSEVR---GKKV-----RAVADIKENREELAKLETLD	107
Qy	158	ATSLASHFIPNVYSTNCKDILHNVAQIAHCYAOKKIGSGPD-----YATA	207
Db	108	TGKTLSEASTADM-----DIHNWEMTFAGLAKXOGGELIINSFIPNAESKVYKEVGAVTQ	162
Qy	202	I----YGLI-VYRRFQPALINDVFOYLESDEPKFPTLEKLIIESNWEKHEKCTLLPYG-I	255
Db	163	ITPMNPYPLQAQSWKILAPALATGCSIMVK-----PSEITPLTTIRVFELMEEVGFPGTII	216
Qy	256	KLLM-----GDVKGSGSETPKXVS	273
Db	217	NLVLAGSSEVGDVMSGHEEVDLV	240

RESULT 9
US-09-495-406-13
; Sequence 13, Application US/09495406

Query Match	4.3%	Score	96.5	DB	4	Length	347
Best Local Similarity	18.2%	Pred. No.	0.82				
Matches	65	Conservative	55	Mismatches	120	Indels	117
						Gaps	13

QY 4 PKOTSJK-----ESRIKISPOFANNGMEYHISNTKEPREVOSRINPF-----84
 Db 43 PKTTLVQINKNNMEDIKLNKAIIGEGHKGF--NYDEKSDKPSPLNPMAFIRVNEAL 100
 QY 85 -LEATFIYLAIVPTFAEDLEIIIIYSD-----PGHSGEDIEETK 123
 Db 101 TLKASLESLPALQ-----RGVIGNDCTDSEIIIEFCKQYPSFRIKRIPIEIOQN 154
 QY 124 TSSNGEFTLVHSRAITEVEKTLGSSAGLVSVATSLSHFTIPNVISTNKDILIHNAQT 183
 Db 155 PSKEENKLSY--NYVAFIP-----KDEWLKIDV 184

```

QY 18 AHCSAOKIGSGEDVATAYGLIVRRRPOALINDVFOVLSDEKPEYELKLIISNWE 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 185 DH1YDAKKLKYSEFYIPAKYIDVVISRVDIHVFNDNPLCDN-----NGNIL 233
QY 244 EKHERCTLEPYGIKLIMGDVYGSGSEYPLKVSRY-LQMKKEPESS----- 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 KEGPDDCLLNNYNNLWKKEY-----LIDIDNNWKKATQSSSSNHSJOLKYKHRI 285
QY 288 VVYDQNSANLQPMKELREMR-EKY-----DSDETYIKELDHSVEPLTVAIKIRK 338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 285 LFFHTELANNHREPLKKHRODIYKKNWISIEBPKFYIQNIMKIGESMISKEYLTK 341

```

```

RESULT 10
US-09-914-259-11
Sequence 11, Application US/09914259
Patent No. 6495336
GENERAL INFORMATION:
APPLICANT: Makowski, Lee
APPLICANT: Hyman, Paul
APPLICANT: Williams, Mark
TITLE OF INVENTION: STRAGED ASSEMBLY OF NANOSTRUCTURES
FILE REFERENCE: 8471-010-999
CURRENT APPLICATION NUMBER: US/09/914,259
CURRENT FILING DATE: 2000-11-21
NUMBER OF SEQ ID NOS: 180
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 3878
TYPE: RPT
ORGANISM: Homo sapiens
US-09-914-259-11

```

Query Match	4.3%;	Score 96;	DB 4;	Length 3878;
Best Local Similarity	20.4%;	Pred. No. 49;		
Matches	88;	Conservative	69;	Mismatches 180;
				Indels 94;
				Gaps 17

```

QY 35 MHAIVITKRSIKSRIRIKISPOFA-----NGEMEYHISNTEKPREVOSRINPPEA 87
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1789 MMSKVTBEGTSLSRLYRSG--FAGTEIDPENBELMINISS-----RLQAAVEKLEA 1833
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 88 TIFIVAYIOPTFAFDLEIIYSDPGYHSDTEKTSNGBKFTLYHSRAITVEKGTGL 147
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1840 -ISETSQLEHAKYOTELM--RESFROKQEATESIKQCEBLRETHIESRAREQALVEL 1899
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 148 GSSAGLVSVAATSLSHFIPNVISTNDILHNAQIAHYA-----QKTISSGPDV 198
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1897 SKAGVTDGYAD- -KTLFERQIOEKTIDILRLOEILCASNRLQELAEQOOIOEBREL 195
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 199 ATAIYGLIYVRFPQALINDVFOVLESDBPEKFPTELKGLISNMBEKHERCTLYPIGKIL 258
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1955 -----LSRQKAMKAEGAPVEQOOLDETEL--MEKLEVCQAEKVRDIOQKQVALL 200
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 259 MGQYKGSSETPKLVSRVLOMKKEKPRESSVVYDOLNSANTOFMKLEBMEKRYSDP--- 315
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 2006 EIDVE-----EQVSRFIELEQEKNTIELM--DILROONQALKEQLEKMRKFLLEQOAI 2055
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 316 -----ETIKELDHSV-----EPLTVAIKNIRKGLQJLTOK--SEVPIBPD--VQ 356
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 2056 EHERDVEQOETIQLEQOIKVPRQPISEHOTREVBOLANHLKEKTDKCELSILSKEOQ 2113
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 357 TOLLDRQOEIPGCVGVVPGAGYDAIVLVLNVOVGNFKQKTLBNPDYFNHVVWDLEE 416
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 2116 RDIOERNBIE-----KLBFVRBELQALLVSADTFQKVE---DR 2155
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 417 QTEGVLEBKE 427
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 2153 KHFQAVAEAKPE 2163
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

```

RESULT 11
US-09-107-532A-5664
; Sequence 5664, Application US/09107532A

```

: Patent No. 6583275
: GENERAL INFORMATION:
: APPLICANT: Lynn A Doucette-Stramm and David Bush
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
: NUMBER OF SEQUENCES: 7310
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: GENOME THERAPEUTICS CORPORATION
: STREET: 100 Beaver Street
: CITY: Waltham
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02354
: COMPUTER READABLE FORM:
: MEDIUM TYPE: CD-ROM ISO9660
: OPERATING SYSTEM: <Unknown>
: SOFTWARE: ASCII
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/107,532A
: FILING DATE: 30-Jun-1998
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/085,598
: FILING DATE: 14 May 1998
: APPLICATION NUMBER: 60/051571
: FILING DATE: July 2, 1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Ariniello, Pamela Deneke
: REGISTRATION NUMBER: 40,489
: REFERENCE/DOCKET NUMBER: GTC-012
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (781) 893-5007
: TELEFAX: (781) 893-8277
: INFORMATION FOR SEQ ID NO: 5664:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 342 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: YES
: ORIGINAL SOURCE:
: ORGANISM: Enterococcus faecium
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (8) LOCATION 1..342
: SEQUENCE DESCRIPTION: SEQ ID NO: 5664:
US-09-107-532A-5664

Query Match 4.3%; Score 95.5; DB 4; Length 342;
Beet Local Similarity 22.2%; Pred. No. 1;
Matches 67; Conservative 45; Mismatches 101; Indels 89; Gaps 14;

OY 196 PDVAT-AYGGLVRRFPALINDVFOLESDEPEKFPTELKK-----LLESNNEE 244
Db 7 YDIPFESHANMDVKRSLOTILLKKAISKER--EKLIDLVKKGLISTEGDLDLESMA TE 64
OY 245 KHERCTLPYGIKL-----LMGDVKGSS-----TPKLVSFVLQWKKEK---- 282
Db 65 KDKQIKKEADKVTASHKEKDQASQIDIKLENGEBEISEPVDPKERERQDQENLEKIDE 124
OY 283 -PESSVVYDQINSANLOF-----MKLEKREKRYDDDEPY---I 319
Db 125 LATANKTSARLDDEVNAQFADNKARNEKQELMQLNTEKEELGTEBELLAQRTLEAEI 184
OY 320 KELHSHV-----PLTVAIKNIRKG-----LQALTOKSEVPIEPVOTOLDRCOEIP 367
Db 185 KELEASGPTLLBEQIKLEAEIKDRIKNQMSKEKEFTTKFELP--DWKKDQATDILNQ-- 240
OY 368 GCVGGVVAGAGYDAIAVLVLNQGNGFKOKTLE--NPDYFNAYVYVVDLBEQTEGVLEEK 425
Db 241 --VEKKNSEAG-----SQGKTLKTKTFQTVSVETVANDNMKQVSLRVPGIATTK 287
OY 426 PE 427

```

```

Db          288 FE 289

RESULT 12
US-09-030-995-3
Sequence 3, Application US/09030995
Patent No. 6235519
GENERAL INFORMATION:
APPLICANT: Mang, Yongzhao
APPLICANT: Childs, John D.
TITLE OF INVENTION: Squires, Charles H.
TITLE OF INVENTION: A Gene Involved in Thlophene
FILE REFERENCE: EBC97-05
CURRENT APPLICATION NUMBER: US/09/030,995
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 520
TYPE: PRT
ORGANISM: Homo sapien
US-09-030-995-3

Query Match          4.3%; Score 95.5; DB 3; Length 520;
Best Local Similarity 20.0%; Pred. No. 2;
Matches 87; Conservative 64; Mismatches 134; Indels 151; Gaps 21

Cy 22 PIYDAYVTKLSS-----KHAAYTPKGTSLKSRKIKIS-----SPQFA-NGEWEYHI 67
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 73 PIPDHYPNFMHNAQVLEYFRMYA--KEFDLLKTYIRKRTYCVSKQDPFATSGQWEVVT 129
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Cy 68 SSWTEKPREVQGRINPFEATIFVIAYIOPTFAFDLEIIISDPG-----YHSOE- 118
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 130 ESEGGKENV--PGQWAVCTGHTNMAL-PLSEF-----PGLKPKKGQYFHSROD 176
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Cy 119 -----DTEKTSNGEKTFLYHSRAITEVEKTG-LSSAGLVS 156
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 177 KNEPGFTGRVVIIGIGNSGDLVAEISQTAQVFLSTRGAMILINRVGDYPA--DV 213
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Cy 157 VATSLLSHITPNVISTNDKILHNAQILNHCQAOKKISSGEVDVATAYIGLVNRR--FQPA 214
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 234 LFSSRLTHPIWKICQS-----LANKYLEKKINQRPD--HEWGLPKRRALSQHP 282
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Cy 215 LINDVQVLESPEKFPTELKGLISNNEBKHERCTLPYGIKLMGVDVKGSETPKLVS 274
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 283 TLND-----DLPRKILS-----GLVKKGVKVKETET---AA 311
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Cy 275 VLOWKKKEDESSVYVDOLNSANLQFMKE-LREMKYDSDPETYIKELDHSVEPLTVAI 333
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 312 IFEDGSRDDIDAVIFATGVSFDPFLSDSVKVVKNKI-----SLYKKVFPNNLERPTLAI 367
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Cy 334 -----KNIRKGIQALTOKSEVPIE-----PDV 355
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 368 IGLIOPGLAIMPISLQGMATQVFKGLTLPSSQSEWMAETSKAOEIDKRYVESORHTI 427
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Cy 356 QFOLNDRCOEIPGCVG 371
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 428 QGDYIDTMELADLVG 443
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 13
US-08-969-046-4
Sequence 4, Application US/08969046B
Patent No. 6455762
GENERAL INFORMATION:
APPLICANT: Chiang, Vincent Lee C.
APPLICANT: Teal, Chung-Jui
APPLICANT: Hu, Wen-Jing
TITLE OF INVENTION: Genetic engineering of trees through
manipulation of lignin biosynthesis
FILE REFERENCE: 881.003051

```


ATTORNEY/AGENT INFORMATION:

FILING DATE: 435
CLASSIFICATION: 435

CLASSIFICATION: 435

US-09-376-330-17
 RESULT 17
 US-09-376-330-17
 Sequence 17, Application US/09376330
 Patent No. 6399321
 GENERAL INFORMATION:
 APPLICANT: Tessier, Daniel C.
 APPLICANT: Dignard, Daniel
 APPLICANT: Bergeron, John J.M.
 APPLICANT: Thomas, David Y.
 TITLE OF INVENTION: Method for screening for
 TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGST) activity
 TITLE OF INVENTION: and nucleic acid encoding for UGST
 FILE REFERENCE: 2139-9"US"
 CURRENT APPLICATION NUMBER: US/09/376,330
 CURRENT FILING DATE: 1999-08-18
 NUMBER OF SEQ ID NOS: 35
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 17
 LENGTH: 1447
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: S. pombe UGST
 US-09-376-330-17

Query Match 4.1%; Score 92.5; DB 4; Length 1447;
Best Local Similarity 20.2%; Pred. No. 21;
Matches 79; Conservative 66; Mismatches 137; Indels 109; Gaps 19;

QY 18 LVLEPIYDAVYVRLSRMAV-----TPKSTLSKESRKIS---SPQ-----FANGE 62
DB 435 ILKPIYPOQLHMLGQHLTVIYPIPPSSPSSPLPILSELIQFRRRPSVQTGMVCAAND 494
QY 63 WE-----YHIS--SNTKEP-REVQSRINPFLATIFIVLAYIQTEAPDLIIYS 110
DB 495 DEQAQVCKSPFPIISGSGTDSALKFLYKCLNSDSSADYSLHEHLPLSEHD----- 547
QY 111 DPGYSQEDTETKTSNGEKTFPL--YHSRAITEVEKTGLSSAGLVSVVATSLSHFIPN 168
DB 548 -----DDTLANLKKDLSSEFDFHMSKSNMWNLIGDSSASEV--YNGRISH----- 595
QY 169 VISTKND-----ILHNAQIAHCYQCKIGSGFDVATAIYGLIYRRPQPALINDVQV 222
DB 596 --DENYDRSMYGIPLFEDIPEVOIAVAEGKISEDDNLDFI-----LRDA--S 638
QY 223 LSSDPEKPT-----ELKKLIESNWEKHEKCTL-----PYGIKILMG-DVKGSE 267
DB 639 LTRNPLVYPSAKSISIDIKRVLENVGSILNHEDILLIGSSNAKTSFWLADPNEKEGLE 698
QY 268 TPQLVSRVLQWKKEKEESSVVDQNSANLQFMKE-----JREMKYDSDPEYIKY- 321
DB 699 ILSLADLISEKND-----ANMLIQGKNHVPPLFAKLSSPKSSKRL 744
QY 322 ---LDHSVEPLVATIKIRKGLQALTKQSEV 349
DB 745 QBLNSSLDPSGVDMDKALKFLKKSRAV 775

RESULT 18

US-08-328-254-6
Sequence 6, Application US/08328254
Patent No. 5710022
GENERAL INFORMATION:
APPLICANT: Zhu, Xueliang
APPLICANT: Lee, Wen-Hwa
TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,254
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/141,239
FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CU 1191
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2482 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-328-254-6

Query Match 4.1%; Score 92.5; DB 1; Length 2482;
Best Local Similarity 21.4%; Pred. No. 51;
Matches 83; Conservative 56; Mismatches 140; Indels 109; Gaps 18;

QY 72 EKPREVQSRINPFLATIFIVLAYIQ---PTBA-FDLIIIIYSDGYSQEDTETKTSN 127
DB 1160 DDKRVESILNEMKLDKSLHQEVQMTKIBACELEKIV-----GELKKNSTL 1210
QY 128 GEKTFY---HSRAITEVEKTGLSSAGLVSVVATSLSHFIPNISTNKDILHNAQIA 184
DB 1211 SEKLETFSCDHQELLQRE-----TSEGLNS---DLEHMDK---SSREDIGDAVAKVN 1258
QY 165 HCYAQKIGSGFDVATAIYGLIYRRPQPALINDVQVLESPEKFPTELKLISSNME 244
DB 1259 DSWKERFL---DVNELSRI---RSERKASIEHEBALYLEADLEVQTE-KLCEKDMEN 1309
QY 245 K-----HERCTL.PYGIKILMGDVKGSETPKLVSRVLQWKKEKPE-----SSVVDQ 292
DB 1310 KQKIVLCBEEISVVTSERNQLRGELDTMSKTTALDQISEKMKETQELSHQSECLHC 1369
QY 293 LNSANLQFMKELREMERKYDSDPEYIKELDHSVEPLTVAIKIRKGLQALTKQSEVPI 352
DB 1370 IQVAEAR-VKEKTELLQTLSSDVSSELDKQTH-----LQEKQLS-----E 1409
QY 353 PDVQQLDRQCEIGCVGVPAGGYDAIAVLNENQVNFK-----QKT 399
DB 1410 KDSQALSLTKCE-----LENQIQALNKEKELVYSESLQAR 1446
QY 400 LENPDYFNHVVWDLDEQTEGVLEKPE 427
DB 1447 LSESDYER---LNVSKALEALVKEGE 1470

RESULT 19

US-08-353-700-1
Sequence 1, Application US/08353700
Patent No. 559919
GENERAL INFORMATION:
APPLICANT: YEN, TIMOTHY J.
APPLICANT: RATNER, JEROME B.
TITLE OF INVENTION: NUCLEIC ACID ENCODING A
TITLE OF INVENTION: TRANSLATIONALLY EXPRESSED KINETOCHORE PROTEIN,
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: DANN, DOREMAN, HERRELL AND SKILLMAN
STREET: 1601 MARKET STREET, SUITE 720
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,700
FILING DATE: 09-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: REED, JANET E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HUMAN
US-08-353-700-1

Query Match 4.1%; Score 92.5; DB 1; Length 3248;
Best Local Similarity 21.4%; Pred. No. 79;
Matches 83; Conservative 56; Mismatches 140; Indels 109; Gaps 18;

QY 72 EKPREVSRINPLFATIFIVLAVIQ---PTEA-PDLEIIIVSDPGVHSGEDTETKTSN 127
DB 1888 DRRKRVESLNMKELDSKHLQEVQMTKIEACIELEKIV-----GELKKENSOL 1938
QY 128 GEKTFPLY---HSRAITEVEKTKGLSSAGLVSVVATSLSHFIPNVI STNKDILHNVQAQA 184
DB 1939 SEKLEVFSCDHQELLQRYE-----TSEGLNS-----DLEMHADK---SSREDIGDVAKVN 1986
QY 185 HCYAOKKIGSGFDVAITAIYGLIVRRFPQALINDVFQVLESPEKPTTELKLIISNMEE 244
DB 1987 DSWKERFPL---DVENEISRI---RSEKASIEHEALYLEADLEVVQTE-KLCLERKDNE 2037
QY 245 K-----HERCTLPYGIKILMGDVKGSETPKLVSRVLQWKKEKPEE---SSVYVDQ 292
DB 2038 KQKRVICLEBELSVTSERNQRLGELDTMSKTTALDQSEKMKETQELSHQSECLHC 2097
QY 293 LNSANLQFMKELREMEKEDSDPEYIKELHDSVEPLTVAIKNIRKGLQALTOKSEVPLE 352
DB 2098 IQVAEAE-VKEKTELLQTLSSDVSSELKDKTH-----LQEKLOSL-----E 2137
QY 353 PDVQVQLDRQCEIPGCVGVPAGGYDAIAVLVLENOVGNFK-----QKT 399
DB 2138 KDSQALSLTKCE-----LENOIAQLNKEKELLVKESSLSQAR 2174
QY 400 LENPDYFNHVVWDLDEQTEGVLEEKPE 427
DB 2175 LSESDYER---LNVSKALEALVKEGE 2198

RESULT 20
PCT-US95-16216-1
Sequence 1, Application PC/TUS9516216
GENERAL INFORMATION:
APPLICANT: Yen, Timothy J.
APPLICANT: Rattner, Jerome B.
TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16216
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,700
FILING DATE: 09-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.

REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 4.1%; Score 92.5; DB 5; Length 3248;
Best Local Similarity 21.4%; Pred. No. 79;
Matches 83; Conservative 56; Mismatches 140; Indels 109; Gaps 18;

QY 72 EKPREVSRINPLFATIFIVLAVIQ---PTEA-PDLEIIIVSDPGVHSGEDTETKTSN 127
DB 1888 DRRKRVESLNMKELDSKHLQEVQMTKIEACIELEKIV-----GELKKENSOL 1938
QY 128 GEKTFPLY---HSRAITEVEKTKGLSSAGLVSVVATSLSHFIPNVI STNKDILHNVQAQA 184
DB 1939 SEKLEVFSCDHQELLQRYE-----TSEGLNS-----DLEMHADK---SSREDIGDVAKVN 1986
QY 185 HCYAOKKIGSGFDVAITAIYGLIVRRFPQALINDVFQVLESPEKPTTELKLIISNMEE 244
DB 1987 DSWKERFPL---DVENEISRI---RSEKASIEHEALYLEADLEVVQTE-KLCLERKDNE 2037
QY 245 K-----HERCTLPYGIKILMGDVKGSETPKLVSRVLQWKKEKPEE---SSVYVDQ 292
DB 2038 KQKRVICLEBELSVTSERNQRLGELDTMSKTTALDQSEKMKETQELSHQSECLHC 2097
QY 293 LNSANLQFMKELREMEKEDSDPEYIKELHDSVEPLTVAIKNIRKGLQALTOKSEVPLE 352
DB 2098 IQVAEAE-VKEKTELLQTLSSDVSSELKDKTH-----LQEKLOSL-----E 2137
QY 353 PDVQVQLDRQCEIPGCVGVPAGGYDAIAVLVLENOVGNFK-----QKT 399
DB 2138 KDSQALSLTKCE-----LENOIAQLNKEKELLVKESSLSQAR 2174
QY 400 LENPDYFNHVVWDLDEQTEGVLEEKPE 427
DB 2175 LSESDYER---LNVSKALEALVKEGE 2198

RESULT 21
US-09-572-191-2
Sequence 2, Application US/09572191
Patent No. 6355466
GENERAL INFORMATION:
APPLICANT: Berard, Christophe
APPLICANT: Sakowicz, Roman
APPLICANT: Wood, Kenneth
TITLE OF INVENTION: No. 6355466el motor proteins and methods for
TITLE OF INVENTION: their use
FILE REFERENCE: 1017
CURRENT APPLICATION NUMBER: US/09/572,191
CURRENT FILING DATE: 2000-05-17
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 1388
TYPE: PRT
ORGANISM: Human
US-09-572-191-2

Query Match 4.1%; Score 92; DB 4; Length 1388;
Best Local Similarity 19.0%; Pred. No. 22;
Matches 59; Conservative 64; Mismatches 111; Indels 76; Gaps 9;

Query Match 4.1%; Score 91.5; DB 4; Length 280;
Best Local Similarity 23.9%; Pred. No. 1.7;
Matches 62; Conservative 32; Mismatches 120; Indels 45; Gaps 10;

QY 109 YSDPGHSDQEDTETKTSSNGEKTFLYHSRAIT-----EVEKTLGSSAGLVSVATSLT- 162
DB YSDIYALGNLHFRSINPDSFGYAKVIALKAEVEINGAKKMLFAQVAYDELIE 97
QY 163 SHPIPNVISTNKDILHVAQIACYAKKIGSGFDVATATYGLIVRRFPALINDVFOV 222
DB LHRGQKLTFTSIEVNPFAVTGEAYL---VGLAFTDTPASLG-----TOIMER 142
QY 223 LRSDEPEKFTPELTKLIESNMEKHERCTLPYGIKLMGVKGSSEPK--LVSRVLQWKK 280
DB ASANPEANPFVKGKODNNLFTAAEADLP-----EDQDAPAKLPFSKVDWLK 193
QY 281 EKREBSVYVDQANSANLQPMKELRENR---KYSDPEYIYELDHSVEPLTVAIKNIR 337
DB 194 POOEQ-----QENQNDQF--KEVADSLFAIAKTGHNQTKLVKVDTSYSELHNKSKLE 246
QY 338 K---GQALQKSEVPIEP 353
DB 247 KERNDLAKLGEENSTP 265

RESULT 25
US-09-107-532A-6075
; Sequence 6075, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6075:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 453 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature

LOCATION: (B) LOCATION 1...453
SEQUENCE DESCRIPTION: SEQ ID NO: 6075:
US-09-107-532A-6075

Query Match 4.1%; Score 91.5; DB 4; Length 453;
Best Local Similarity 18.1%; Pred. No. 3.8;
Matches 69; Conservative 78; Mismatches 143; Indels 91; Gaps 16;

QY 92 VLAVIQPTAFPLEIITIIYSDPGHSDQEDTETKTSSNGEKTFLYHSRAITVEKTLGSSA 151
DB LIAFVE--NOAMENPLVEVEPEMOPDYIKASSSTGEET-----NYLAQIDPTGSLFD 99
QY 152 GLSVVATSLSHPIPNVISTNKDI--LHVAQIACYAKKIGSGFDVATATY---GLI 206
DB SLIEQVHLYNDTFLRKIVLYVEYIDLNGFLITISLEAGKEIG-----ATPIQMLDLIT 154
QY 207 VYRRFPALIN---DVFVLESDEBKPTEL-----KKLIESNMEKHERCTLP- 252
DB 155 LIQOLEPAGVARSLOECIMLQTERDDYAFELAYIVLECEFELEVRKKKEIAQRPVDL 214
QY 253 YGKILMGDKGSETP--KLVSRLQWK-----KEKREBSVYVDQANSANLQPMKE 303
DB 215 HAVQOIFDYLQTLSPSPGRIFPSSSEFNIPVRVLDNDDKNVQVISRKQNPNIIRFQES 274
QY 304 -LRENEKYSDPEYIYELDHSVEPLTVAIKNIRK----- 339
DB 275 YFKQSQQADKETENVYLERKQEFELKKTII--LQRPDILHVAQIVSRQAPFLDKER 332
QY 340 -IQALTQKSEVPIEDVQTLDRCEIIPCGVGVPAAGYDAIAVLVLEQVGNFKOK 398
DB 333 PIKPLTLK--EVATIEDVHSIVSR-----AVNGKLTEDFGVFELK 372
QY 399 TLENDPYFNHYVVDLEOTE 419
DB 373 -----KFTTRIPINSTEOTE 388

RESULT 26
US-09-134-001C-4378
; Sequence 4378, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4378
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4378

Query Match 4.1%; Score 91.5; DB 4; Length 878;
Best Local Similarity 18.6%; Pred. No. 11;
Matches 91; Conservative 90; Mismatches 195; Indels 113; Gaps 21;

QY 6 SAPGRFLAGVLYLEPIYDAVVTALSSRMHAVITPKGTSLKESRIKI-----SSPQ 57
DB 260 SSIDKIKISMGYL-----DNLSLDLSSNLNLTIGRGIG--KSTLIELRVALDIAPTSQ 312
QY 58 FANGEMEYHISNTEKPREVOSRINPFEATIFVLAVIQPTAFPLEIITIIYSDPGHSDQ 117
DB 313 NNTSPFNCKSNLIGGKVE-----LITSHAQIGKQPKIIRKNYEDP---II 358
QY 118 EDTETKTSNGEKTFLYHSRAITVEKTLGSSAGLVSVATSLSHPI-----PN 168

```

Db      359 KIDNNVSNVTKDILPNIENVSONEIIIDLTNNEN---AKNIIINFLDKDRSDNKK 414
Qy      169 VISTN-----KDIHNVAQ-----IACVAAKKIGSGFDVATAIYGLI 206
      415 EIKTNHNSKSLIKAKEDLENQEKINOLPKAKELKH-FNLGLIGCKLEAVOGKISRE 473
Qy      207 VYRFPQALIND---VFOVLESDEPKFTEK-----KLISSNMEKHERCTLPGY 254
      474 QYIQTOKQIIBNDISITNIIIFENENYNOQIGVIEIPDSIKITDNHKKLKE----- 527
Qy      255 IKLMGVDYGGSEPTLVSRVLQ-WKKEKPESSVYVDOLNSANLOPMELNEMRKYS 313
      528 ILSMFTDLK-DTQNEIKRTIYVWKKKN-----TEKINRA-----IKSLDIEGKTE 577
Qy      314 D---PEYIKELDHSVEPTVAIKNIKGLQALTOKEVPIBPVOTOLLDRCOEIPQCV 370
      578 DIAHETTEROKQITSIETLEPTQLSRKTSIETL-ENHRQLKEDLKEIREDQKLNLRVY 636
Qy      371 G-----GVPGAGGYDAIAVLLENQVGNFKOKTLENPDYFNHYWYDL--E 415
      637 KKIINRYLKQVNIKIOPYANVNLIIEFLKEENGLDSTLKWIKNHQSFPNPKFIKID 696
Qy      416 EOTEGVLEE 424
      697 RDESEALYEE 705

```

RESULT 27

```

US-09-198-452A-171
; Sequence 171, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffeis, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, previe
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 171
; LENGTH: 1156
; TYPE: PRF
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-171

```

Query Match 4.1%; Score 91; DB 4; Length 1156;

Best Local Similarity 18.8%; Pred. No. 20; Matches 75; Conservative 56; Mismatches 147; Indels 122; Gaps 16;

```

Qy      23 IYDAVYVALSS-----RMHAVITPKGTSLKESRIKISPPQANGEMEHYISSNTE--- 72
      764 LIBEAVILNVTASQGLCRHQMQA-----SFR-RILINPDGAKHGBAERTLASREMLKT 815
Qy      73 -----KREVOGRINPFLIATIFVLAVIOPTFAFDEIIIYSDPGYHSQED 119
      816 IGLSLTLPFVRSSPESQSGYNQLKVR-----EQLFDE-----QRQONET 859
Qy      120 TETKSSNGEKTFLYHSRAITE--VEKTGLSSAGLVSVVATSLIS--HFLPNVI----- 170
      860 VSPPEYAVAOALAAVVRGHESLIIVTYGLAQEOGQTSKVTTLMRDLAAVAVELVEMVE 919
Qy      171 -----STNKDILHNVAQIAHCYAOKKIGSG--FDVATVIGLIYVRRPOPALINDV 223
      920 TYRLNRSDOILHRVSVLHSHLRDSDSGNGIIVVKKLFELINNGNNP----- 969
Qy      224 ESDPEKFTLEKLIESNMEKHERCTLPGYIKLMGDVKGSETPKLVSRVLQWKKKEP 283
      970 -NDPFC-----QKYMQIILDAPVSLLYGAFFSKF----- 997
Qy      284 EESSVYVVDOLNSANLOPMELNEMREKYSDP-----ETIYIKELDHSVEPTVAIKNIK 339
      998 NEFLINFTELINIAN--STLAAEBAKRYVEEKGRTYEWBEAKORLEIAIAELDDLHNO 1055

```

```

Qy      340 IQALTOK-----SEVPIBPVOTOLLDRCOEIPG 368
      1056 ETLLEQIRIANLNLKISIFSDNLNREKVSVEKALBEEIIG 1095

```

RESULT 28

```

US-08-062-368-2
; Sequence 2, Application us/08062368
; Patent No. 5491086
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Wang, Alice M.
; TITLE OF INVENTION: Purified Thermostable Nucleic Acid
; TITLE OF INVENTION: Polymerases Enzyme From Pyrodictum Species
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingstland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/062,368
; FILING DATE: 19930514
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sias Ph.D., Stacey R.
; REGISTRATION NUMBER: 32,630
; REFERENCE/DOCKET NUMBER: 8584
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2863
; TELEFAX: (510) 814-2977
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 803 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-062-368-2

```

Query Match 4.1%; Score 90.5; DB 1; Length 803;

Best Local Similarity 20.7%; Pred. No. 12; Matches 99; Conservative 64; Mismatches 197; Indels 119; Gaps 24;

```

Qy      20 LEPIYDAVYVALSSRMHAVITPKGTSLKESRIKIS---SPQANGEMEHYISSNTEKPR 75
      113 LEGVEDSLEADIRRAMRYLIDKLYPFTAYVRAENAGRSGP-RVDSVYITVEDPEPIA 171
Qy      76 EVQGRINPFLIATIFVLAVIQ--PREAPLEIITIIYSDPGYHSQEDTETSSNGEKT 132
      172 DITSIDIPKEMVLAFDIEVYKSGSPNPSRPVLI-----SIDSKGNEKXL 219
Qy      133 ---LYHSRAITEVEKGTGLSSAGLVSVVATSLSHFIPNVISTNKDIL--HN----- 179
      220 EANNYDDR-----NVLEFTEIYISPPDIIIVGINSNNFMPY 257
Qy      180 VAQIAHCYAOKKIGSGPDV-----ATAIYGLI-YVRRPOPALINDV-----FOVL 223
      258 LIBRAH-----RIGVKLDVTRVGAEPSSMSYGVAGRLNVDLYVVEEMHEIKVXTL 312
Qy      224 ESDPEKFTLEK---LIE-----SNMEKHERCTLPGYIKLMGDVKGSETPKLVSRV 275
      313 EEVAEVLGVWKEKSRVLEIEMWRIPDYWDDEKR--PLKKRYALDDVRA--TYGLAKKI 366
Qy      276 LQWKKERPEESSVYVDOLNSANLOPMELNEMREKYSDP--ETIYIKELDHSVEPTVA-- 331

```

Db 367 LPAIQLSTVTVGPDLQVGMVGFLWTLKRAADHMLVBNRYKREESYKGAIVLK 426
Qy 332 AINIKKGLQALTKQSEVP-----IBPVQQLDRCOEIPGCVGVPAG-----G 379
Db 427 PLKGVHENVVLPSSNYPNIMIKYVNGPDTIIDPSECEKSGCY--VAPEVGHMFRS 484
Qy 380 YDAIAVLVLNVOG-----NFKOKLENDPYFHNVYVWDLSEOTGVLEKPEYIG 431
Db 485 PSGFPTVLENLIALKQVKEKKEPPSPSPER--RIY--DERQKALKVLANASYGMG 539

RESULT 29
US-09-198-452A-1030
; Sequence 1030, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1030
; LENGTH: 945
; TYPE: PRF
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-1030

Query Match 4.1%; Score 90.5; DB 4; Length 945;
Best Local Similarity 21.3%; Pred. No. 16;
Matches 82; Conservative 46; Mismatches 110; Indels 147; Gaps 23;

Qy 4 AFSAPGKAFIAGYVLEPIYDAVYVVALSRMAVITPK-----GTS 45
Db 460 AFTYPMNKT-----VFVFSVSHSAFSDALDQVHLFINPKRQEDLREKYAVHQBPAHP 514
Qy 46 LKESR-----IKISSPQ-----FANG-----EW-EVHIS----- 68
Db 515 LSDGRVVRHRIQULVAPGHCARFGCGNASTLPTVTKEKAEWFKLHSENNCAIAYTS 574
Qy 69 ---SNTKE-----PREVQ--SRINPFLA-----TIFVLAVIOPTBAFDEIIT 108
Db 575 APLSKAKQSKIFSQIPKSKYERQEPFLPSGDTSLKLYINQA--IQTSMLDIYMH 633
Qy 109 Y--SDP-----GYHSQEDTETKTSNGEFTLYHSRAITEVE---KTGLSSAGLVSVVA 158
Db 634 YESHPIPLGCTYALAEVLNENSKSLVSLKNEQILTDVDFEFPSSINTGFEYISYEL 693
Qy 159 TS-----LASHPIPNV-----ISTNKDILANVAQIACVAK----- 190
Db 694 TEKGRKHYSCVIDSTQYALRYIOEHGIPNYTLBEIST-----INALNVCYSKSPLPD 746
Qy 191 -----KISGPDVATATGILVYRRF-----OPALINDVOYLESDEPEFTLEKLTIE 239
Db 747 LKQVLSLQNE--DLSTPYHSLVTPKYSSEDESALIN-----LVSDPEQARVLSKNS 800
Qy 240 SNWEK---HERC--TLPGIKLMG 260
Db 801 EHMEBATQLHDPIFDMTYVVKALDG 825

RESULT 30
US-09-107-532A-6343
; Sequence 6343, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneka
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-8277
; TELEFAX: (781)893-5007
; INFORMATION FOR SEQ ID NO: 6343:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 823 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...823
; SEQUENCE DESCRIPTION: SEQ ID NO: 6343:
US-09-107-532A-6343

Query Match 4.0%; Score 90; DB 4; Length 823;
Best Local Similarity 19.2%; Pred. No. 14;
Matches 87; Conservative 72; Mismatches 164; Indels 130; Gaps 20;

Qy 8 PGKAFIAGYVLEPIYDAVYVVALSRMAVITPKGSLKSRKI-----SSPQFANG 61
Db 219 PGDPFTGGILQKQKIKKAYETG-----RGKVLRSKTKIEMKGNKQOIVIT 267
Qy 62 EMEYHSSNTEKPREVQSRINPFLAETIFVLAVIOPTBAFDEIITSDPGVHSGEDTE 121
Db 268 EIPYENKATLVKQDEIRLANKKIDG---INAVRDESDRTGLQIVV-----E 311
Qy 122 TKTSSN--GEKFTLYHSRAITEVEKTLGSSAGLVSVVATSLASHPIPNVISTNKDILAN 179
Db 312 LKKNDAAGILNVLFPN---TELQ---INYNMVAI-----DHMTPOQVGL--KDIAS 358
Qy 180 VAOIACHVAKKIGSGFDVATATGILVYRRQPAL-----IN 217
Db 359 YIEHRQVITKR--SOFDLAKAQKQHOIVEGLMKALSLIDEVIATIRSKDKKAKKILV 416
Qy 218 DVFOVLESDEPEKPT--ELKGLIESNWEKHERCTLPYGIKLMGDVKGSETPKVSRVL 276
Db 417 DVFOFTEQAEAVITQLRLNRTDITELQK-----EESLIAQIT 457
Qy 277 QWKK--EKPESSVYVQQLNSANLQFMKELEMEKRYDSDEFTYIKELDHSVPLTVAI 333
Db 458 ELNKILSNKELFSVW-----KKELEBVKKVNSASRLTIED-----EI 496
Qy 334 KIRKGLQALTKQSEVP-----EPDVQQLDRCOEIPGCVGVPAGYDAI-- 383
Db 497 ERIKIDQVLVAQEDVIYVTVREGYVKTSLRSYSASKEEIGMEBGDYLVSGLSTLD 556

QY 384 AVLLENQVGNFKQKLENDYFHNWVWDLLE 416
| : : : : :
DB 557 HVLITNKANVIRPVHELDP-----LKMWDAGE 585

RESULT 31
US-09-328-352-4409
; Sequence 4409, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4409
; LENGTH: 244
; TYPE: PRN
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4409

Query Match 4.0%; Score 89.5; DB 4; Length 244;
Best Local Similarity 25.7%; Pred. No. 2.2;
Matches 44; Conservative 30; Mismatches 50; Indels 47; Gaps 11;

QY 223 LESDEKPFTEKLIIESNWEKHERCTIPYGIKLMGVKGSFPLKVSRYLQMKKEK 282
| : : : : :
DB 50 LIDDDPLAMQJANEIQA-YSDK-----PF---ILFGHSVGGG---LIMKVLWYLNK 95

QY 283 PESSVYVQDLSANIQFMKE---LREMEKYDSDETYKELD-----HSEVPL 329
| : : : : :
DB 96 P-----LIDQLLVISSPEHHYIQHMYKIELDEKIDILKXNNPDEILNNQEL 150

QY 330 TVALNIRKGL-----QALTO---KSEVPI-----EPDV-OTOLLRCQE 365
| : : : : :
DB 151 TFFLKIRNDPFLSDQLSEKIHKTEVPIVTFYQKDPIDPKRWMDAMQ 201

RESULT 32
US-09-257-825B-21
; Sequence 21, Application US/09257825B
; Patent No. 6403352
; GENERAL INFORMATION:
; APPLICANT: Poovaiiah, Bachettira W.
; APPLICANT: Patil, Shameekumar
; APPLICANT: Takezawa, Daiuke
; TITLE OF INVENTION: Compositions and Methods for Production of Male-Sterile Plants
; FILE REFERENCE: 4630-51993
; CURRENT APPLICATION NUMBER: US/09/257, 825B
; CURRENT FILING DATE: 1999-02-25
; PRIOR APPLICATION NUMBER: US 08/655,352
; PRIOR FILING DATE: 1996-05-23
; PRIOR APPLICATION NUMBER: US 60/014,743
; PRIOR FILING DATE: 1996-03-28
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 517
; TYPE: PRN
; ORGANISM: Nicotiana tabacum
US-09-257-825B-21

Query Match 4.0%; Score 89; DB 4; Length 517;
Best Local Similarity 18.0%; Pred. No. 8.3;
Matches 68; Conservative 69; Mismatches 118; Indels 122; Gaps 17;

QY 14 AGGYVLE-----PIYD-----AYVALSRMNAVITTPGTSIAKSRITIS 54
| : : : : :
DB 112 SGVHLILECGSGGLFDRLAGARYNEMGAADVVOIAKGLAL---HGASIVHRDLKPE 168

QY 55 SPQFANG-----WEYHISNTEKPREVQSRINPFLPATIFIVLAYIQPTAEDLEIT 107
| : : : : :
DB 169 NCLFLNKENSPKIMDFGLS-----IEDFANPVV--GLFGSIDVSP-EAL----- 213

QY 108 IYSDPGYHSDTEKTSNGEKTFLYHRAITEVEKGLGSSAGLVSVATLSHFIP 167
| : : : : :
DB 214 -----SRENTYTK-----SDWSLGVITLYILSGVP 240

QY 168 NVISTNKDILHNVAIAHCAQKIGSG--FDVATAIYLIVRFPALINDVFOYLE 224
| : : : : :
DB 241 FLAPENR-----KQMLNQPFSDEKT-----WGNISSAKOLISLILK 281

QY 225 SDPEKPFTEKLIIESNWEKHERCTIPYGIKLMGV-KGSETPKLVSRYLQMKKEP 283
| : : : : :
DB 282 VDPNMRPT-AQELIEHPW-----VTGLAKQOMDABIVSRLOQSFNSRK 325

QY 284 EESSVYVQDLSANIQFMKELEMEKRYDSDETYKELDSVEPL-----TVALNIRK 338
| : : : : :
DB 326 PRAAMASVLSFSLRTKJKLVGSYDLKPEE-LQWLSHNFKKICNGENSTILLEPEE 384

QY 339 GLQALTOKSEVPIEPDV 355
| : : : : :
DB 385 VKAMENSLVPLAPRI 401

RESULT 33
US-09-134-001C-5153
; Sequence 5153, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5153
; LENGTH: 304
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5153

Query Match 4.0%; Score 88.5; DB 4; Length 304;
Best Local Similarity 19.8%; Pred. No. 3.9;
Matches 65; Conservative 55; Mismatches 122; Indels 87; Gaps 14;

QY 45 SLK--ESRIKISSQFANGEMWEYHISNTEKPREVQSRINPFLPATIFIVLAY-IQPTEA 101
| : : : : :
DB 35 SLKNDKDALIIVSLP--FNASW-----LNRYSIMQALIIDVIGYEVKPHFI 79

QY 102 FDLRIIYSDPGYHSDTEKTSNGEKT-----FLYHRAITEVEKTS---LGSSAG 152
| : : : : :
DB 80 SEDLASINNVNTQSEVQEPVOHSSIDDKTGKQOFNMANTFDFVIGPGRFPFAASLA 139

QY 153 LVSVAATSLSHFIPNVISTNKDILHNVAIAHCAQKIGSGPDVATAIYGLIVYRRFQ 212
| : : : : :
DB 140 VAAAPAEAYNPLFIYGVGLGK--THLMHAIQHVLNKKPNA----- 179

QY 213 PALINDVFOYLESPEKPFTEKLIIESN---WEKHERCTIPYGIKLMGVKGSSET 268
| : : : : :
DB 180 -----KVITYTSSEKTFNFIKSRINDTEAREKRYKIDV-----LLIDDI----- 220

QY 269 PKLVSRYLQMKKEPPESSVYVQDLSANIQFM-----KELREMEKRYSDPE-----T 317
| : : : : :
DB 221 -----QPLQNKQOEPEFHTFNLHQNNKOIYVSSDRPPEIAKLEBRLASRFEMGLIV 275

QY 318 YIKELDSVEPLVVALNIRKGLQALTOK 346
| : : : : :
DB 318 YIKELDSVEPLVVALNIRKGLQALTOK 346

Db 276 DITPPDYKTR-MAILQKLLKILIFROK 303

RESULT 34

US-09-328-352-4877
; Sequence 4877, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4877
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4877

Query Match 4.0%; Score 88.5; DB 4; Length 1005;

Best Local Similarity 26.3%; Pred. No. 28;

Matches 36; Conservative 23; Mismatches 47; Indels 31; Gaps 7;

QY 322 LDHSEVPLTVAIKNIRK--GLQALTKQSEVPIEPDV-OTQLDRCOE-----FGCV-370

Db 1 LNSAPHYITSTKILKKGSKMVIYIGTPKQDITQPOVLEKVKRIEVDNAGVVR 60

QY 371 -----GGVVGAGGYDAIAVL-----VLENGVGNFKQTLLENPDYFHNVYVWDL 415

Db 61 STTMSLAARQAKGIEANAGFPAKLLPSSSVTQEDIDHLKLLALNPTMNSV--VSXD 118

QY 416 EOTEGV---LEEREDY 429

Db 119 QQTAAILELLESPSGP 135

RESULT 35

US-09-328-352-5102
; Sequence 5102, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5102
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5102

Query Match 3.9%; Score 88; DB 4; Length 437;

Best Local Similarity 24.4%; Pred. No. 7.8;

Matches 38; Conservative 33; Mismatches 57; Indels 28; Gaps 9;

QY 21 EPIYDAYVVALSSRMHNV-----ITPKGTLKESRIKISSPPA--NGEMEHYISSNTER 73

Db 183 EYVRAVYFEGGKXKHAVDLRLTR---LQOQKLLTGNPIFLNIHIERHLINDA 239

QY 74 PREVQSRINPFLFAT-----FIYLVAYIQP--TEAFLEIITISDPGYHSOE---DT 120

Db 240 LNEIQA-LHPELDVFTALDPMWMLLAIYKELIAYVDIQLKVY-PLSYGRDPMWML 297

QY 121 ETKTSNGEKTFLYHSRAITEVKTGLGSSAGIVSV 156

Db 298 ATRVSKRTVAFTPPCR---PTKESTLGAQKLFTYSV 330

RESULT 36

US-09-425-383-2
; Sequence 2, Application US/09425383
; Patent No. 6194637
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; APPLICANT: Shi, Jinhui
; TITLE OF INVENTION: Maize DNA Ligase I Orthologue and Uses
; FILE REFERENCE: 0962
; CURRENT APPLICATION NUMBER: US/09/425,383
; CURRENT FILING DATE: 1999-10-22
; EARLIER APPLICATION NUMBER: 60/108,793
; EARLIER FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 909
; TYPE: PRT
; ORGANISM: Zea mays
US-09-425-383-2

Query Match 3.9%; Score 88; DB 3; Length 909;

Best Local Similarity 21.2%; Pred. No. 26;

Matches 76; Conservative 55; Mismatches 141; Indels 86; Gaps 16;

QY 55 SPQFANGEMEHYISSNTER-----PREVQSRINPFLFATFIYLVAYIOPTAFDLEIITY 109

Db 235 SPTKAKQ---ASQSEKRAAPKAKTTDSPTEKNTLEL--KKGSEFDPMAVAY 288

QY 110 SDPGYHSQEDTETKTSNGEKTFLYHSRAITEVKTGLGSSAGIVSVATSLSHPIPV 169

Db 289 WKPG-----EPVFFLARALDLI-----SNESG-RIVITTEILSNVPRV 327

QY 170 ISTN-KDILANVAQAACHCAQKIGSGFDVATATYGLVYTRFQALINDVQVLESDE 228

Db 328 MATPDDLLATVYLSANRIAPHEGIEIGIDA-----SVIRALAEVGRKE 375

QY 223 KPTEKLLIESNMEKEHRCITLPGIKILMDVKGSETPLVGRVQMKKEPDESSV 288

Db 376 HYKDLKEIGDGLVAKASRSQKMFK-----PKPLTIRAVLSTFTTAKESG- 424

QY 289 VYDQNSANLQPKELREMRKY-----DSDETYIKELDHSVEPLVAIKNIRKG 339

Db 425 -----KQSDKKRHHMGLVAATDCPQ-YITRLQSKRIGLAEKTVQWA 470

QY 340 L-QALTKQSEVPIEPDVOTQLDRCOEIPGCVGIVPAGGVDALAVLVLENGENK 396

Db 471 LGQAAVYADKXPLPKVQSPFEBAKIKQAY-SVLP---IYDKIVPALR--VGWVK 522

RESULT 37

US-09-107-532A-5256
; Sequence 5256, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5256:
SEQUENCE CHARACTERISTICS:
LENGTH: 1114 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...1114
SEQUENCE DESCRIPTION: SEQ ID NO: 5256:
US-09-107-532A-5256
Query Match 3.9%; Score 88; DB 4; Length 1114;
Best Local Similarity 17.4%; Pred. No. 37;
Matches 76; Conservative 67; Mismatches 186; Indels 108; Gaps 14;
QY 35 MAVAITPKGTSKSEKRIKISSPO---FANGEMEHISNTEKREPVOSRINPLEATIF 90
DB 554 LAAVRRVEGEYERLKEIPILNDAKTLTLFQGETSGVFQESAGINVAIRLDPENIEDIA 613
QY 91 IYLAIVOPTFEAFDLIIISDPGYS---QEDTETKTSNGEKFELYSRAITVEKTL 147
DB 614 AVNALYRPGPMQNTIFLARKKSKAIRPDSLPIKNTYGVIVYQEQ-IMQASTMA 672
QY 148 GSSAGLVSVATSL-----LSHFIPNVISTNDIINVAQIAHCAQKKIGSGF 196
DB 673 GFTLGSODILRAVSKKKVDLEEKSHVSGAL--QQYSEKTAATVYEIEKFRANPF 730
QY 197 DVATAI-YGLIYRRFPOLINDVQVLESDEKPTTELKILIESNWEKHERCTLPYGI 255
DB 731 NRSHPAFVSFV---GFQMAVYK-----AHMGAFETSLMNSVRHN----- 767
QY 256 KLIMGDVKGSETPKLVSRVLQMKKEKPESSVYVDOLNSA-----NLQFMREL 304
DB 768 -----TSKLKEVIARAKNKKLSSPSINQSSPELVNKEIRGLTAIRGI 815
QY 305 RE-----MREKYSDPEYTIKELDSVEPLTVAIKIRGLQALTKSEVPIDEPVOTQ 358
DB 816 RRDVEEDILNERKQNGP-----FQSVQFLIRIDKRWLKLLEQLSLAVGVFDELSPN 868
QY 359 LLDRCQELPGCGVGPAGYDAVAIVLENQVGNFKOKTLENDYFHNVTWDLSECT 418
DB 869 RKQMLDLGKIQNTIVSGGSLDLGLGALKE-----EEVA 904
QY 419 EGVLEEK---PEDYIGL 432
DB 905 DYTLEERLKEBEYIGI 921
RESULT 38
US-09-252-991A-21217
Sequence 21217, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 21217
LENGTH: 181
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21217
Query Match 3.9%; Score 87.5; DB 4; Length 181;
Best Local Similarity 23.9%; Pred. No. 2.1;
Matches 48; Conservative 36; Mismatches 70; Indels 47; Gaps 10;
QY 186 CYAQKKIGSGFVATAIYGLIV-----YRFPOLINDVFOVLESD-----PEKFP 231
DB 7 CFVQECYRKFTQFVLTALMAASAPAEKIAIVANQMALLESDAKQYAVDAKFTG 66
QY 232 TELKKLIESNWEKHERCTLPYGIKLLMGDVKGSETPKLVSRVLQMKKEKPESSVYVD 291
DB 67 PQLNLT-----KNLERDAKALQD--KLVSNGSKMGSGDERKALDPK 106
QY 292 QLSANLQPM-KELREMEKXDSPEYTIK-ELDSVPEPLVAIYNIRKG--LQALTKQS 347
DB 107 Q-KADFPQSKELNESAAADRMDLKKLKRKLQDAV-----ETIKKGGYDNIERGA 159
QY 348 EVPIEP--DVQTLDRCEI 366
DB 160 VDVKFPQYDITRQYIERMQL 180
RESULT 39
US-09-107-532A-3929
Sequence 3929, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3929:

SEQUENCE CHARACTERISTICS:
LENGTH: 258 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...258
SEQUENCE DESCRIPTION: SEQ ID NO: 3929;
US-09-107-532A-3929

Query Match 3.9%; Score 87.5; DB 4; Length 258;
Best Local Similarity 19.1%; Pred. No. 3.7; Indels 63; Gaps 11;
Matches 50; Conservative 44; Mismatches 105; Indels 63; Gaps 11;
QY 97 QPTAFDLEIIISDPGYSQEDTEKTSNGEKTEFLYHSRAITEVEKGTGSSAGLVSV 156
DB 3 QPTKFCSS-----YMFENDASTCSRTKGTSPFPRKPAISVNDHFTKGSTPMITIV 54
QY 157 VATS--LISHFIPNVISTNKDILHNAVQIAHCYAKKIGSGFDVAIAIYGLIYRRFP 213
DB 55 IIDDDPFVTKSLQTIIESTKEI--RVLGIGHC-----AKDALVYETHRPD 98
QY 214 ALINDFVQVLESDE-----KEPTLKKLIESNMEKHERCTLPYGIK-----L 257
DB 99 VLTIDIRMKQOTIDAKELAFPKAILLITTPDEETIRKAFSGIKGYIKONLQA 158
QY 258 LMGDKG-----GSETPKLVSRVLQWKEKEPSSVVYDQANSANIQPMKEL----- 304
DB 159 IIPSVASVANGQAVGNE---IIEFTFQWKKNKPTIDSHVAFSRELAIIEVAAGKN 215
QY 305 -REMRKY--DSDPEYIKEL 322
DB 216 NKEIADALYLSDETGVNYSQL 237

RESULT 40
US-09-643-597-369
Sequence 369, Application US/09643597
Patent No. 6426072
GENERAL INFORMATION:
APPLICANT: Wang, Tongrong
APPLICANT: Pan, Liguun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Hosken, Nancy
APPLICANT: Fanger, Gary R.
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Henderson, Robert A.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.455C11
CURRENT APPLICATION NUMBER: US/09/643,597
NUMBER OF SEQ ID NOS: 369
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 369
LENGTH: 708
TYPE: PRT
ORGANISM: Homo sapiens
US-09-643-597-369

Query Match 3.9%; Score 87.5; DB 4; Length 708;
Best Local Similarity 17.3%; Pred. No. 19; Indels 125; Gaps 22;
Matches 78; Conservative 64; Mismatches 164; Indels 125; Gaps 22;

QY 28 VTALSRMAVITPKGTSIKESRIKISSPQFANGEMHYHSSNTEKREVSQSRINPLEA 87

DB 110 VCHLGDQLEGVNTPPQRAVEAQKMKYENEFJDLGKSPVFTNSEKIKKAADIIQK----- 165
QY 88 TIFIVLAIYQPTAFDLEIIISDPGYSQEDTEKTSNGEKTEFLYHSRAITEVEKGTG 147
DB 166 --LHLIAQELPPDRF-----SEVKKIKSK-----YHDLQCQLIQFTS 202
QY 148 GSSAGLVSVV--ATSLISHFIPNVISTNKDILHNAV--AQIAHC-----YAKKIGSGFDVAT 200
DB 203 AQRGEISRMREVAVALHF-----KGYHCVDVYIKQCEGAYLRNDI---FEDA- 250
QY 201 AIYGLIYRRFPQPALINDVQVLESDEPEKPTLKKLIESNME-----EKHERCTLP 252
DB 251 ---GILCQRVNRQ--VGDI-----SNPE---TVLAKLIQNVFEIKLQSFVKEQLSECRKS 298
QY 253 YGIKLMDGVKSGSETPKLVSRVLQWKEKEPSSVVYDQANSANIQPMKELREMRKYD 312
DB 299 DAEQYIKNLIDYITNTLSSKLMENL-GTDKQFPLSKLISFISYENYIEVETGYL 357
QY 313 SDP-----ETVYIKELDSVEPL-TVAIKIRKGLQALQKSEVPIEPDVOT----- 357
DB 358 KSRSAMLIQRYVDSKNHOKRSIGTGIDLK--ERIRQRTNLPFGPSIDTHTGTFPLSGE 414
QY 358 -----QLDRC-----QEIPEGVGVVPAGGYDAIATLV-----LEN 390
DB 415 VVNNLQETKQAFERCHRLSDPSDLP-----RNAFRIFTIIVEFLCIBHIDYALET 465
QY 391 QVGNFKQKLTENPDYFHNWYWDLEQTEGV 421
DB 466 GLAGIPSSDSRNA---NLFFLDVVOQANTT 492

Search completed: August 8, 2003, 09:41:06
Job time : 34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2003, 09:34:51 ; Search time 41 Seconds
(without alignments)
103.290 Million cell updates/sec

Title: US-10-069-062-7

Perfect score: 2230
Sequence: 1 MSKAPGKAFKFLAGGYLV.....DLREQTEGVLEKPEYDYLGL 432

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	809.5	36.3	451	2	phosphomevalonate
2	611.5	27.4	426	2	probable phosphome
3	475	21.3	505	2	unknown protein (I
4	225	10.1	323	2	(phospho) mevalona
5	173	7.8	358	2	phosphomevalonate
6	170.5	7.6	359	2	mevalonate kinases
7	165.5	7.4	360	2	AEI434
8	129	5.8	317	2	F70185
9	126	5.7	404	1	SE2440
10	124.5	5.6	443	1	BVBYR1
11	123	5.5	378	1	S42088
12	122.5	5.5	3187	2	JCS837
13	120	5.4	449	2	G81419
14	119.5	5.4	526	1	PSKR15
15	117	5.2	335	2	D97914
16	117	5.2	862	2	E95044
17	116	5.2	335	2	T07775
18	113	5.1	833	2	T32289
19	112.5	5.0	526	1	PSKRBT
20	111.5	5.0	324	2	F72474
21	111	5.0	1249	2	H71404
22	111	5.0	4092	1	S38128
23	110.5	5.0	600	2	C69371
24	110.5	5.0	721	2	T41942
25	110.5	5.0	790	2	T22849
26	110.5	5.0	829	2	F89630
27	110	4.9	664	2	C71106
28	109.5	4.9	643	2	B75055
29	109.5	4.9	986	2	T52176

30	109.5	4.9	1332	2	D84669	aldehyde oxidase I
31	108.5	4.9	1778	2	JT0382	apolipoprotein B -
32	108.5	4.9	2663	1	S28261	centromere protein
33	108	4.8	260	2	B70419	hypothetical prote
34	108	4.8	792	2	G86564	phenylalanyl tRNA
35	108	4.8	792	2	D81608	phenylalanyl tRNA
36	108	4.8	792	2	H72058	phenylalanyl tRNA
37	107.5	4.8	743	2	A29232	101K malaria anti
38	107.5	4.8	4450	2	JX0340	gramicidin S synth
39	107	4.8	1790	2	S67593	transport protein
40	107	4.8	2166	2	G70163	hypothetical prote
41	107	4.8	2954	2	T14156	kinesin-related pr
42	106.5	4.8	646	2	A82428	hypothetical prote
43	106.5	4.8	897	2	C90561	hypothetical prote
44	106	4.8	1161	2	H95903	probable chemotaxi
45	105.5	4.7	844	2	AC1981	hypothetical prote

ALIGNMENTS

RESULT 1

S57588 phosphomevalonate kinase (EC 2.7.4.2) - Yeast (Saccharomyces cerevisiae)

N:Alternate names: protein YW9595.02; protein YMR220W

C:Species: Saccharomyces cerevisiae

C>Date: 19-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000

C:Accession: S57588; A39606

R:Skellton, J.; Churcher, C.M.

submitted to the EMBL Data Library, June 1995

A:Reference number: S57587

A:Accession: S57588

A:Molecule type: DNA

A:Residues: 1-451 <SKE>

A:Cross-references: EMBL:Z49939; NID:G887599; PID:G887601; MIPS:YMR220W

A:Experimental source: strain AB972

R:Tsay, Y.H.; Robinson, G.W.

Mol. Cell. Biol. 11, 620-631, 1991

A:Title: Cloning and characterization of ERG8, an essential gene of Saccharomyces cerevi

A:Reference number: A39606; MUID:91117228; PMID:1846667

A:Accession: A39606

A:Molecule type: DNA

A:Residues: 1-212, 'R', 214-417, 'PLMT', 422, 'D', 424 <TSA>

A:Cross-references: EMBL:M63648; NID:G553127; PIDN:AAA4596.1; PID:G171479

C:Genetics:

A:Gene: SGD:ERG8

A:Cross-references: SGD:S0004833; MIPS:YMR220W

A:Map position: 13R

C:Keywords: phosphotransferase

Query Match	Score	36.3%	Score	809.5	DB 2	Length	451
Best Local Similarity	41.2%	Pred. No.	4.9e-45				
Matches	190	Conservative	76	Mismatches	146	Indels	49
Gaps							14
QY	3	KAFAPGKAFKFLAGGYLVLEPIYDAYVYVTAISRMHAAVTPKQTSIKES---	RIKISSPQFA	59			
DB	5	RAFAAPGKALLAGGYLVUDTTEAFVGLSARMHAAVHPYV-SIQSGSKFEVRYKQPK	63				
QY	60	NGEMWHISSNTE-KPRBVQSRINPFLFATIFIVLAYIQPT-----EAFDLBITIYSD	111				
DB	64	DGEMLYHSPKSGFIPVSIKSKNPFIEKVIANFVSFKPMDDYCNRMNLFVID--	121				
QY	112	PGYHSOEDTERKTSNNEKFTLYHSRAITEYEKTKGLSSAGLVSVVATSLSHFI--	168				
DB	122	DAYHSQSDSVTE--HRNNRRLSFHSRIEVPKTKGLSSAGLVTVLTAAALSPVSDLEN	179				
QY	169	VISTNKDILHNVAQIAHCYAOKKIGSGFDVATAIYGLIVYRRFPQALINDVFOYLESPE	228				
DB	180	NVDKXREYVHNLAQVAHQOQKIGSGFDVAAAAYGIRRRFPALISNIPDI---GSA	236				
QY	229	KFPPELKKLI-FSNWEKHEKCTLPYGIKILMGDVKGSGSETPKIVSRVLOKKKEPSS	287				
DB	237	TYGSKLAHLVDEEDMNTIKSNHLPISGLITLMGDIKNGSETVKLVQKVKMNYDSHMPESL	296				

submitted to GenBank, April 2001

A:Description: *Sulfolobus solfataricus* complete genome.

A:Reference number: A99139

A:Accession: F90479

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-323 <KUR>

A:Cross-References: GB:AE006641; NID:g13816374; PIDN:AAK43093.1; GSPDB:GN00155

C:Genetics:

A:Gene: SSO2988

Query Match 10.1%; Score 225; DB 2; Length 323;
Best Local Similarity 23.7%; Pred. No. 2.7e-08;
Matches 100; Conservative 67; Mismatches 141; Indels 114; Gaps 16;

QY 6 SAPKAFLAGYLVLEPIYDAVVTALSSRMHVAITPCKGSLKSRKIKSSPPFANGEMWY 65
DB 5 SAPKILMTGYSV-----FGISHVIAVANKRVSCSLREIKKOSLIFFTSYGH 54
QY 66 HISSNTEKREVQSRINPFLKATIFVLAVIOTPEAFDELIITYSDPGYHSQEDTETKTS 125
DB 55 PKNSGNEL---INSVLDTFRRRLSQLPOCY-----EIDLND----- 88
QY 126 SNGEYTFLYHSRAITEVEKTKGLSSAGLVSVVATSLSHFIDPVISTNKDI--LHNVAOI 183
DB 89 ---KEF-----IIDGKKTGLSSSS---AATVSLTFLCTLYAIGHKLDLFEIHKLAOI 133
QY 184 AHCAOKKIGSGFDPVATATYGLIYRRFPQPALINOVQVLBSPDKPTELKLIJESWE 243
DB 134 ANVKOKIGSGFDPVATATYGLIYRRFPQPALINOVQVLBSPDKPTELKLIJESWE 180
QY 244 EKHERCTLPYGIKLMGDVKGSETPKLVSRVLQWKEKREBSVVYDQLSANLQFMKE 303
DB 181 ---MGLGFTGKSSFTVGLVRKFE-----KSNLDDPKE 210
QY 304 LREMEKTDSPETIKELHDSVEPLTVAIKIRKGLQALTOXSE---VPIEPVQTL 359
DB 211 IMRL-----IDENYMAIKLIMLKLDEAVEHIKGRKTLNIAIRIVGVKLVSKHEBEL 265
QY 360 LDRCEIPGCVGVPGAGVDAIVLVLENQVGNFKQKTMENPFIHNVAVDLEBQTE 419
DB 266 IKTAE-EGALVALSPGAGGDSI--FALGNLNRREMSK---RGIFLIDVKEE-E 316
QY 420 GV 421
DB 317 GL 318

RESULT 5

B89828 phosphomevalonate kinase [imported] - *Staphylococcus aureus* (strain N315)

C:Species: *Staphylococcus aureus*

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 12-May-2003

C:Accession: B89828

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc

ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, R.; Inoue, R.; Kaito, C.; Sekimizu, K.;

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: B89828

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-358 <KUR>

A:Cross-References: GB:BA000018; PID:g13700484; PIDN:BBB41781.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: mvak2

C:Superfamily: mevalonate kinase

Query Match 7.8%; Score 173; DB 2; Length 358;
Best Local Similarity 22.8%; Pred. No. 0.00014;
Matches 97; Conservative 57; Mismatches 143; Indels 128; Gaps 20;

QY 7 APKAFLAGYLVLEPIYDAVVTALSSRMHVAITP---KGT-----SLKES 49
DB 6 APKLYINGEYAVTEPGKSVLILMDRVATITEADQYKGIHKKALHNHNVFSSRED 65
QY 50 RIKISSPPFANGEMWYHISNTEKREVQSRINPFLKATIFVLAVIOTPEAFDELIITY 109
DB 66 SIVSDPFAAAQ-----LNVVVAIAIEFEQYAKSCDIAM--- 99
QY 110 SDPGYHSQEDTETKTSNGEYTFLYHSRAITEVEKTKGLSSAG--LVSVAVATSLSHFIPN 168
DB 100 ---KPHHLTIDSNL--DSSNG-----HRTYGSSAAVAVSVI--KVLNEFDIM 140
QY 169 VISTNKDILHNVAOIAHCAOKKIGSGFDPVATATYGLIYRRFPQPALINOVQVLBSP 227
DB 141 KLSNL--YIVGLAVIANKQSLSSCG--DIASVYSGLAIVSTDHEVHQIIE----- 191
QY 228 EKFPTELKLIJESWEEKH--ECTLPGIKLMGDVKGSETPKLVSRVLQWKEKREBS 286
DB 192 ---DTVEEVLIKWPGLIHPELOAPENMEVLIQWTSPPASPHVSEVRKLL----- 241
QY 287 SVYDQLSANLQFMKELEKREKTDSPETY---IKELDSVEPLTVAIK--NIRKGLQA 342
DB 242 ---SDPSFYGDPLEBDSHRCVBEKLHAFTNNIKGVOK 275
QY 343 LTOKEVPIE-----PDVOTO---LDRCEIPGCVGVPGAGVDAIVLVLENQ 391
DB 276 MVRQNRITIQMDKEATYDITETKLYCDIAEKYH--ASKTSGAGGDC--GITIINKD 332
QY 392 VGNFK 396
DB 333 VDKEK 337

RESULT 6

AE1806 mevalonate kinases homolog lmo0012 [imported] - *Listeria monocytogenes* (strain EGD-e)

C:Species: *Listeria monocytogenes*

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 12-May-2003

C:Accession: AE1806

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker

.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.

D.; Jones, L.M.; Karet, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapk, G.; Madueno, E.; Maitouram, A.; Ma

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,

A:Title: Comparative genomics of *Listeria* species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AE1806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-359 <GLA>

A:Cross-References: GB:NC_003210; PIDN:CAC98227.1; PID:g16409371; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lmo0012

C:Superfamily: mevalonate kinase

Query Match 7.6%; Score 170.5; DB 2; Length 359;
Best Local Similarity 24.7%; Pred. No. 0.00021;
Matches 96; Conservative 55; Mismatches 144; Indels 93; Gaps 19;

QY 8 PGKAFLAGYLVLEPIYDAVVTALSSRMHVAITPCKGSLKSRKIKSSPPFANGEMWYH 67
DB 13 PGKLYINGEYAVTEPGKSVLILMDRVATITEADQYKGIHKKALHNHNVFSSRED 65
QY 68 SSNTEKREVQSRINP-----PLEATIFVLAVIOTPEAFDELIITYSDPGYHSQED 118
DB 58 ENPVSWE--VGGELKPPGHEMTPTAEAINATITLK--SEGELLPVPMII----- 104
QY 119 DTEKTSNGEYTFLYHSRAITEVEKTKGLSSAGLVSVVATSLSHFIDPVISTNKDILH 178
DB 105 ETELIDSGA-----KYGLGSSAAATVAIVNALMTKFPISIMLKK---F 146

```

QY 179 NVAQIAHCAOAKKIGSGFVATAIY-GLIVRRFPALINDVQVLESDEPEKTEKTL 237
DB 147 KLAALSHLVQGN-GSCGDIASCWGWTATTFDEBWK--HRLAYKSLFMKEMPM 203
QY 238 ISSNMEKHERCTLPYGIKLMGVKGSSETP---KLVSRLQWKKEKPESSVVDL 293
DB 204 LQI---ETLEBPVPTFSV-----GWTGTPVSTGKLVSQIHAFK---QEDSKNYQH 248
QY 294 NSANLOFMKEIREMKRYSDPEYIKELDHSEPLTVAIKIRKGLQALTKSEVPIEP 353
DB 249 LTRNNEIMQIIOAFHTKOE-----ELLYSAIKENRRILQELGTAGVNIET 295
QY 354 DVQTOILDRCOEIPGCVGVPGAGGYD 381
DB 296 SLIKELADSAENNGG--AGKSSGSGGD 321

RESULT 7
AB1434
Mevalonate kinases homolog lin0012 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 12-May-2003
C:Accession: AB1434
R:Glaser, P.; Frangul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.
Science 294, 849-852, 2001
A:Authors: Krefel, U.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitouran, A.; Me,
ok, C.; Schleuter, T.; Simoes, N.; Tixeret, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AB1434
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-360 <GLA>
A:Cross-references: GB:AL592022; PIDN:CA95245.1; PID:g16412433; GSPDB:GN00178
A:Experimental source: strain Clip11262
A:Genetics:
A:Gene: lin0012
A:Superfamily: mevalonate kinase

Query Match 7.4%; Score 165.5; DB 2; Length 360;
Best Local Similarity 24.0%; Pred. No. 0.00047;
Matches 93; Conservative 58; Mismatches 144; Indels 93; Gaps 19;

QY 8 PEKAPLAGYLVLEPIYDAYVTALSSRMNAVTPKGTSLKSRIKISSQFANGEMETI 67
DB 14 PGLVYAGAYAVESGHTAILTA---VNRYYT---LTLEDSB-----RRLMIPHY 58
QY 68 SSNTEKPREVQSRINF-----FLKATIFVLAYIOPTAFDL---ELIYSDPGVHSG 118
DB 59 ENPVSPF--IGGHLKDGDEHWTFTALAINATTFELK--SGCIELTPVAKMYI----- 105
QY 119 DETKTSNGEKTFLYHSRAITEVEKTLGSSAGLVSVAATLSLHFIPIVISTNEDILH 178
DB 106 ETELIDQSGA-----KYGSSAAATVAVINMLMTKFPYELISLTK--F 147
QY 179 NVAQIAHCAOAKKIGSGFVATAIY-GLIVRRFPALINDVQVLESDEPEKTEKTL 237
DB 148 KLAALSHLVQGN-GSCGDIASCWGWTATTFDEBWK--HRLAYKSLFMKEMPM 204
QY 238 ISSNMEKHERCTLPYGIKLMGVKGSSETP---KLVSRLQWKKEKPESSVVDL 293
DB 205 LQI---ETLEBPVPTFSV-----GWTGTPVSTGKLVSQIHAFK---QEDSKNYQH 249
QY 294 NSANLOFMKEIREMKRYSDPEYIKELDHSEPLTVAIKIRKGLQALTKSEVPIEP 353
DB 250 LTRNNEIMQIIOAFHTKOE-----ELLYSAIKENRRILQELGTAGVNIET 296
QY 354 DVQTOILDRCOEIPGCVGVPGAGGYD 381
DB 297 SLIKELADSAENNGG--AGKSSGSGGD 322

```

```

RESULT 8
P70185
Phosphomevalonate kinase homolog - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: F70185
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White,
son, D.; Peterson, J.; Kevlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
, Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, R.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: F70185
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-317 <KLB>
A:Cross-references: GB:AE001169; GB:AE000783; NID:g2688611; PIDN:AAC67030.1; PID:g268861
A:Experimental source: strain B31

Query Match 5.8%; Score 129; DB 2; Length 317;
Best Local Similarity 22.1%; Pred. No. 0.14;
Matches 93; Conservative 52; Mismatches 126; Indels 150; Gaps 22;

QY 4 AFSAGKAFIAGYLVLEPIYDAYVTALSSRMNAVTPKGTSLKSRIKISSPOF---A 59
DB 5 SFSVGNLLMGEYITL-----EKGLGL--ATAIKRAAFPSFK 42
QY 60 NGEWEYHISNTEKPRE---VQSRINPLEATIFVLAYIOPTAFDLIILYSDPGYHS 116
DB 43 SDSMRF--SKKKKIDDSLLENRSD-----FVFGMFAVLSQNCPEFNLENFAY----- 88
QY 117 QEDTETKSSNGEKTFLYHSRAITEVEKTLGSSAGL--VSVAATLSLHFIPIVISTKMD 175
DB 89 --DAVIDSN---FFPN---DGTGKGSSVAVAIGVCGFLIHMTNV--BK 134
QY 176 ILHNAQIAHCAOAKKIGSGFVATAIYGLIV-----YRRFPALINDVQVLES 226
DB 135 EIFKYLEARYSOGGISTGDIATISIRGVIEPBGFNPKCRQIANGVDFYLMQGLQ 194
QY 227 PEKPETELKILSNMEKHERCTLPYGIKLMGVKGSSETPKLVSRVLQWKKEKPEBS 286
DB 195 AIKTT---SICEVN---KHNSILDPIIKCNL-----EMKCLV----- 227
QY 287 SVVYQALNSANLQ--FMKEIREMKRYSDPEYIKELDHSEPLTVAIKIRKGLQALT 344
DB 228 -----LNANSKSKALISLRRAK-----LGLAIG-----EAIG 256
QY 345 QKSEVPIEPDVQTOILDRCOEIPGCVGVPGAG-----GYDAIAVLVLN 390
DB 257 VSAALPSSFD--HLLGQCDILKAL-----GAGNETFLVTRPNIENRNLKIIISIVLEN 307
QY 391 Q 391
DB 308 E 308

RESULT 9
S62440
Mevalonate kinase (EC 2.7.1.36) - fission yeast (Schizosaccharomyces pombe)
N:Alternate names: hypothetical protein SPAC136.11c
C:Species: Schizosaccharomyces pombe
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Dec-1999
C:Accession: S62440; 137646
R:Odell, C.; Bowman, S.
submitted to the EMBL Data Library, October 1995
A:Reference number: S62430
A:Accession: S62440
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-404 <ODE>

```


A:Accession number: Z21734
A:Reference: 137646
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-404 <OD2>
A:Cross-references: EMBL.Z54308, PTDN.CAA91104.1; GSPDB:GNO0066; SPDB:SPAC13G6.11d
A:Experimental source: strain 972h-, cosmid c13G6
C:Genetics:
A:Gene: SPDB:SPAC13G6.11c
A:Map position: 1L
A:introns: 25/3
C:Superfamily: mevalonate kinase
C:Keywords: ATP; isoprenoid biosynthesis; phosphotransferase

Query Match	5.7%;	Score 126;	DB 1;	Length 404;
Best Local Similarity	24.6%;	Pred. No. 0.33;		
Matches 81;	Conservative 41;	Mismatches 121;	Indels 86;	Gaps 14

```

QY 1 MSKAF--SAGCKFFLAGGYLVEIPIDAVYVLTSSRMHAIYIPKOTSLKERIRISSP- 56
Db 1 MSKSLIVSPGKITLGEHAV--VYGAFALAAVSISSYCKLTQTTNNELVIMSDIGT 57
QY 57 ----QFANGWEHHSINTKPE-----REVQSTINFELEATFIYLAY 95
Db 58 ERKRNLTOSLPHQHVAVENVOHPASSPNLDLLOGLBELLKNEBNGLIHSAMCTLYPTSL 117
QY 96 IOETFAEDLEIIYYSDPGYHSODEPTKTSNGEKTFLYHSRAITFEVKTGLSSAGLVS 155
Db 118 SSPSGCCTLTIISSQVPLG-----PALGSSAITSV 146
QY 156 VVATSLISHPIPVISTKNDILHNAQIHCYQKKIG-----SGFDVATVY-GLI 206
Db 147 VVATSLILAA-GNIEPPSSNSLONNKALILBAMFLEBCCIHGTPSGIDNAVANTGGLI 205
QY 207 VYER---FQPAL-----INDVFOYLESDEPEKTEPLAKLIESNMEKHERCTLYGKIL 258
Db 206 AFRKATAQSAAMEFLPKDITLSVMITD-TKQPKSTKLVGVGFELK-ER--LPTVIDSI 261
QY 259 MGDVKGSGSEPKLVSVLVOMKKEKEBESS 287
Db 262 IDAIDIGISK---SAVLALTSESDDNSS 285

```

RESULT 10

mevlonate kinase (EC 2.7.1.36) - Yeast (*Saccharomyces cerevisiae*)
N:Alternate names: protein YMR261.02; protein YMR208w
C:Species: *Saccharomyces cerevisiae*
C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 21-Jul-2000
A:Accession: S05875; #S1531; S55090
R:Kearsey, S.E.; Edwards, J.
Mol. Gen. Genet. 210, 509-517, 1987
A:Title: Mutations that increase the mitotic stability of minichromosomes in yeast: characterization
A:Reference number: S05875; MUID:86121728; PMID:332847
A:Accession: S05875
A:Molecule type: DNA
A:Residues: 1-443 <KEA>
A:Cross-references: EMBL:X06114; NID:g4286; PIDN:CAA29487.1; PID:g4287
R:Oulmouden, A.; Karst, F.
Curr. Genet. 19, 9-14, 1991
A:Title: Nucleotide sequence of the ERG12 gene of *Saccharomyces cerevisiae* encoding mevalonate
A:Reference number: S1531; MUID:91243209; PMID:1645230
A:Accession: S1531
A:Molecule type: DNA
A:Residues: 1-443 <OUL>
A:Cross-references: EMBL:X55875; NID:g3683; PIDN:CAA39359.1; PID:g3684
R:Dedman, K.; Brown, D.; Bowman, S.
submitted to the EMBL Data Library, June 1995
A:Reference number: S55089
A:Accession: S55090

A;Molecule type: DNA
A;Residues: 1-443 <DED>
A;Cross-references: EMBL:Z49809; NID:G854459; PIDD:CAA89923.1; PID:G854460; GSPDB:GN0001
A;Experimental source: strain AB972

A/Gene: SGD:ERG12; PAR1; MIPS:YMR208w
A/Cross-references: SGD:S0004821; MIPS:YMR208w
A/Map position: 13R

C; FUNCTION:

A.Description: catalyzes the reversible phosphorylation of (R)-mevalonic acid by ATP to A.Pathway: isoprenoid biosynthesis

A/Note: this enzyme can also utilize GTP, CTP, and UTP
C; Superfamily: mevalonate kinase

Cineyolous: A1F1 isoprenoid biosynthesis; phosphotransferase
F;348-357/Region: ATP binding #status predicted
F;375/Binding site: ATP (lys) #status predicted

Query Match	5.6%	Score 124.5;	DB 1;	Length 443;
Best Local Similarity	22.2%;	Pred. No. 0.48;		
Matches 105; Conservative	72;	Mismatches 154;	Indels 143;	Gaps 27;

```

QY 1 MSKAF--SAPGAFPLAGGLV--EPIDAVYTLSSMAHVITPKGTSKESIKISSP 56
D 1 MSLPFLTSPGKVIITFGHSAVYKPKVAASVSL--RTYLIS--ESSAPDTEIDFP 55
QY 57 QFA-NGEWEYHIS-----NTEKPREVGRINPFEATIFIVLAYIOPTAEADIEII 108
D 56 DISFNHKSINDFNMATTEDOVSSQLAKAQ-----QATDGLSGLVS 97
QY 109 YSPD-----GHSQEDTETKSSNGETFLY-----HSRATVEKT-----G 146
D 98 LDPPLLAQSESFTHA-----ACFLYMPVCLCPHAKNIKIFSLKSTPLPGAG 145
QY 147 LGSAGLVSVAVTSLSLHSRIPNVISTN--KQILHNAQIAHCYA--QKKI--GSGFDV 198
D 146 LSSSASI--SVGLALAMAYIGSLIGSNDLEKLSENDKHIVQMFAIEBKCHGPGSIDN 203
QY 199 ATAIYS--LIYRRRQPALIN--DVFQVLESDEKFPTELKLLIESNWEKHERCTLPYGI 255
D 204 AAYATYGNALLFEKDSHGNTINTNNEKFLDDPP-----AI 237
QY 256 KLIMGDVGKSGSEPTLVSRYLQMKKEK--PESSVYVQDLSANLQ-----FMKELREMR-- 308
D 238 PMILYITTRIPRSTKQLVARVRLYTEKFEPEMKIILDMGECALQGLEINTKLSKCKGTD 237
QY 309 -EKYDSDEPTY--IKELHSVEPLTVAIKNIKRGIALTKOSEVLEPDDVOTLDRCOE 365
D 298 DEAVEITNNELYEQLLLEIRINHGLLVSIGVSHPGLELIKULSD-----DLR----- 343
QY 366 IGGCVGV--VPAGGYDAIVL-----VLNONGKFKQTELENDPYFHNVMWDL 414
D 344 ----IGSTKLTAGGGGCSLTLRRDITQEOBIDSFKKLQD--DFTYETETD 391

```

RESULT 11

mevalonate kinase (EC 2.7.1.36) - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #ext_change 10-Sep-1999
 C/Accession: S42088
 R/Riou, C.; Tourte, Y.; Lacroste, F.; Karst, F.
 A/Accession: S42088
 A/Reference number: S42088
 A/Accession: S42088
 A/Molecule type: mRNA
 A/Residues: 1-378 <RIO>
 A/Cross-references: EMBL:X77793; NID:g456613; PIDN:CAAS4820.1; PID:g456614
 C/Function:
 A/Description: catalyzes the reversible phosphorylation of (R)-mevalonic acid by ATP to ;
 A/Pathway: isoprenoid biosynthesis
 A/Note: this enzyme can also utilize GTP, CTP, and UTP
 A/Superfamily: mevalonate kinase

C:Keywords: ATP; Isoprenoid biosynthesis; phosphotransferase
F:330-339/Region: ATP binding #status predicted
F:352/Binding site: ATP (lys) #status predicted

Query Match 5.5%; Score 123; DB 1; Length 378;
Best Local Similarity 20.1%; Pred. No. 0.48;
Matches 88; Conservative 74; Mismatches 145; Indels 130; Gaps 17;

```

QY 7 AAGKALAGGYLVLEIYDAYVATLSSRMNAVIT---PKGTSLKESRIKISSQFANGW 63
   |||||
DB 7 APEKTLIAEHAH---VHGSTVAAAIIDYTVTLRPLPSANNDRLTTL---QLKDIDL 60
   |||||
QY 64 EYHISNTEKPREVOGRINPFLERATFIYLAAYIQTEAPDLIIISDPGYSOE----- 118
   |||||
DB 61 EFSWS-----LARIKENIPDSSITLCHSTPASCSEETLKSI 96
   |||||
QY 119 -----DTEKTISSNGEKTFLY-HSRAT-----TEVE-KTGLGSSAGLVSV 157
   |||||
DB 97 AVLVEBQNLPEKEMWSSGISSTFLMLYTRIIIGFNPAVVINSLEPYSGISGSAALCVAL 156
   |||||
QY 158 ATSLSHFI-----DNVISTNKDIIHNVQAQIHCAQKKI-----GSGPDVATAIYG 204
   |||||
DB 157 TALLASSISSEKTRNGMSSLDFTNLNLKWA-----FEGEKTIHGKPSGIDNTVSAYG 211
   |||||
QY 205 LLYVRRFOPALINDVFOVLESDEPEKPTTELKLISSNWEKHERCTLPGYIKLMDVKG 264
   |||||
DB 212 NMI-----KFCSGEITRLQSN-----MP--LMLINTNV 239
   |||||
QY 265 GSETPYLVSRLVQWKKEKPESSVYVDQNSANLQFMKELREMRERYSDPEYIKELDH 324
   |||||
DB 240 GRNTKALVSGVQRAVRHPDAMKSVFNADVISKEIPLAIIQSDFTSVTEKEERIKELME 299
   |||||
QY 325 SYEPLTVALIKNRKGLQALTKQSEVIEPDVQQLDRQCEIRPGCVGVPPAGGVDATA 384
   |||||
DB 300 MNQGLLSLVGSHSSIEAV-----ILTTVKRKLVS-----LTVGGGGGCVL 341
   |||||
QY 385 VLVLNQVNFQKXTLE 401
   |||||
DB 342 TLL---PTGTVDKXVE 355
   |||||

```

RESULT 12

364k Golgi complex-associated protein - rat
JC5837
C:Species: Rattus norvegicus (Norway rat)
C>Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #ext_change 20-Jun-2000
C:Accession: JCS837
R:Tok, C.; Fujiwara, T.; Sohma, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein rec
A:Reference number: JCS837; MUID:98093490; PMID:9431462
A:Accession: JCS837
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3187 <TK>
A:Cross-references: DDBJ:D25543; NID:9516825; PIDD:BA05026.1; PID:9516826
C:Comment: This protein plays a role in the formation and maintenance of the characteris
C:Superfamily: granlin
F:49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper. #status predict
F:3165-3187/Domain: membrane anchor #status predicted <MAD>

Query Match 5.5%; Score 122.5; DB 2; Length 3187;
Best Local Similarity 17.5%; Pred. No. 13;
Matches 80; Conservative 70; Mismatches 155; Indels 153; Gaps 13;

```

QY 28 VTALSSRMNAVITPKGTSLKESRIKISSPOFANGEMEYHISNTEKPREVOGRINPFLRA 87
   |||||
DB 1373 IKALHTQLMOKAKHEHERLKQVQVELCELKQPKLEESKAKQOQRLQALSRKA 1432
   |||||
QY 88 TTFIVLAAYQTEAPDLIIISDPGYH---SOEDTEKTS-SNGEKTFLVHSRAITEVE 143
   |||||
DB 1433 -----LKENKSIOEQLSASARDAVEHRLTKSLADVESQVSVQNCBKDALLGKALLQEE 1484
   |||||

```

QY 144 KTGLGSSAGLVSVATSLSHFIPNVISTNKDIIHNVQAQIHCAQKKIGSGPDVATAIY 203
DB 1465 RKL-----IYEMDSLLENQSLGSCSELKALG----- 1515

QY 204 GLIYVRRFOPALINDVFOVLESDEPEKPTTELK-----KLIS-SNWEKHERCTLPGYIKL 257
DB 1516 -----LTEDKEKIMKELESVCRSKIASTEQHGKHELOKEY--EV 1554

QY 258 LMGDVKGSETPKLVSRVLQWKKEKPESSVYVDQNSANL-----QFMKEL 304
DB 1555 LIQSYEVNSNEARLQHVAVSVROKQ-----VYALRLASBDKREKREKQLODAQOEMEM 1611

QY 305 REMREKYSDPEYIKELD-----HSEPLTVAI 333
DB 1612 KEMKRFPAKSKQOKLLELEBENDRLARAQPYGANGESMEALLSSNASLKELEKILLEY 1671

QY 334 KNIRKGLQALTKQSEVIEPDVQQLDRQCEIRPGCVGVPPAGGVDAIYVLNQV 393
DB 1672 KTLSEKFPALMAEKNTLSEETRNMLQVEADEL----- 1704

QY 394 NFKOKTLENPDYFNHVVWVLEBQTEGVLEEKPEYIG 431
DB 1705 --KQASLET-----TEKSDPKDVIIEVTEAVVG 1731

RESULT 13

hypothetical protein Cj0036 [imported] - Campylobacter jejuni (strain NCTC 11168)
G81419
C:Species: Campylobacter jejuni
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #ext_change 03-Jun-2002
C:Accession: G81419

R:Parhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Baaham, D.; Chillin
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; Vanvliet, A.; Whitehead, S.; Barrel
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: G81419
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-449 <PAR>
A:Cross-references: GB:AL139074; GB:AL111168; NID:96967505; PIDD:CA872529.1; PID:9696754
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj0036

Query Match 5.4%; Score 120; DB 2; Length 449;
Best Local Similarity 22.0%; Pred. No. 1;
Matches 64; Conservative 45; Mismatches 94; Indels 88; Gaps 12;

QY 193 GSGPDVATAIYGLIYVRRFOPAL-----IND----- 218
DB 16 GSFIDISRLVLAQILDAKQKQKKEFDEEVNAKRAEYTKALNDLTKQKIEQKLINE 75

QY 219 -VFQVLESDEPEKPTTEL-----KLISSNWEKHERCTLPGYI 255
DB 76 QVSQQLLEKQFQELLQKQNFQKSEKFNKHEMNMKIMQBELKSKSELSPLSI 135

QY 256 KL-----LMGDVKGSETPKLVSRVLQWKKEKPESSVYVDQNSANLQFMKELREMRERY 311
DB 136 KAENRRLKREOKEMERLKFQAKBAFKEFKQESKNLEPFRKRRLEFQKSTQBDQKY 195

QY 312 DSDPEYIKELDHSVEPLTVAIKNRKGLQALT-OKSEVPIEPDVQQL-DRQCEIRPGC 369
DB 196 -KELETRKSVAQKLED---KORRIEQSQOLQGEAALLIBEYIQSEYLSDBYKVEP-- 249

QY 370 VGVVPPGAGGYDAIYVLNQVGNF-----KOKTLENPDYFNHVVWV 413
DB 250 -----KGVNAGDCLH--IVKDNFGNIGSIIYESRTKE-----FNKEMLD 288

RESULT 14

outer capsid protein VP5 - bluetongue virus (serotype 1, strain Australia)
P5XR15

C:Species: bluetongue virus
 C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #ext_change 23-Jul-1999
 C:Accession: J50323
 R:Gould, A.R.; Pritchard, L.I.
 Virus Res. 9, 285-292, 1988
 A:Title: The complete nucleotide sequence of the outer coat protein, VP5, of the Austral.
 A:Reference number: J50323; M0ID:88237575; PMID:2837015
 A:Accession: J50323
 A:Molecule type: genomic RNA
 A:Residues: 1-526 <GOU>
 A:Cross-references: GB:M21845; MTD:9323193; PIDN:AAA42849.1; PID:9323194
 C:Genetics:
 A:Map position: segment 6
 C:Superfamily: bluetongue virus outer capsid protein VPS
 C:Keywords: capsid protein

Query Match 5.4%; Score 119.5; DB 1; Length 526;
 Best Local Similarity 27.4%; Pred. No. 1.4;
 Matches 75; Conservative 34; Mismatches 114; Indels 51; Gaps 14;

162 LSHF---IPNVISTN--KDIHNVAGIACHYAOCKIGSGFDVATAYGLI----- 206
 8 LSRFGKGVNALTSTAKIYGTIGKAAERFAESIGS-----AALDGLVQSGVHSILTG 62
 207 -VYRRP--OPALINDVFOVLESDEPKP-----TELK-KLIESNWEKHERCTLPYGI 255
 63 ESYGEVYKQAVLLN-----VLGSG-BEIPDPLSGETIEQAKIRELEDEGRNELVRLKYN 117
 256 KILMGDVKGGSTPKLVSVVLQWKEKEPESSVYVDQLNSANLQFMKELREKREKYSDDP 315
 118 KI---KEKGEERLE-VVEFNAGAAVAEVEDERKOPDILNKAVTSYKILTEBEDLQMRRLA 173
 316 EYIEKLDHVSPEPLTVAIINIRKGLQALQKSEVPEPVDVQQLDRCOIEGCGV-- 373
 174 NALQKEIGERTIAETVWVEYRNKIDAL--KNALIEBERGQCE--EALQELAGMTGVLE 229
 374 -----VPGAGYDATALVLENOVGNFKOKTLEN 402
 230 AAEEVPLIGAGMATATVATGATRAIEGAYKLKXVIN 263

RESULT 15

phosphomevalonate kinase [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #ext_change 12-May-2003
 C:Accession: D97914
 R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Bursgett, S.; Dehoff, B.S.; H
 e, R.; LeBlanc, D.J.; Lee, L.N.; Leffkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaekunas, S.R.;
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; M0ID:21429245; PMID:11544234
 A:Accession: D97914
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-335 <KUR>
 A:Cross-references: GB:AE007317; PIDN:AAK9144.1; PID:915457898; GSPDB:GN00174
 C:Genetics:
 A:Gene: mvak2
 C:Superfamily: mevalonate kinase

Query Match 5.2%; Score 117; DB 2; Length 335;
 Best Local Similarity 21.6%; Pred. No. 1.1;
 Matches 92; Conservative 54; Mismatches 147; Indels 132; Gaps 20;

9 GKAFIAGGYLVLEPIYDAVVTALSSRMHAVITPKGTSIKESRIKISSPOFANGMEYHI 68
 8 GKLWAGGYALIEFGQALIKDIPYMAEIA-----FSD--STIKY 47
 69 SNTKEPREVOSRINPLEATITIVLAYIOPTFAFLEIIIVSDPGYHSGEDTETKTS-- 126

DB 48 S-----DMPFAVDLRPNPDYSLIOETIALMGDFL 77
 QY 127 --NGEK-----TFVHSRAITTEVEKTKGSSAGLSVVAATSLISHIPVISTNKOILNV 180
 DB 78 AVRGONLPPFSIATYGRKERGKRGGLSSGVVAVKALALAY--NLSDONILFPL 134
 QY 181 AQIAHCYAOCKIGSGFDVA-TAYGLIYVRRFOPALINDVFOVLESDEKPEFTEKLE 239
 DB 135 TS-AVLKRGNGSGMDLACIAEDLVLYQSFDR--QKVAAMLEE-----NLATYLE 184
 QY 240 SNWEKHERCTLPYGIKILMGDVKGGSTPKLVSVVLQWKEKEPESSVYVDQLNSANLQ 299
 DB 185 RDW-----GFSI--SQVKPTLECDPLVG---WTRKENVVSHWVOQIKONINQN 227
 QY 300 FMKELREKREKYSDDPEYIK-----ELDHVSPEPLTVAIINIRKGLQALQKSEVPI 351
 DB 228 FL-----TSKKEIVVSLVEALEGKSKKIIIOVEVSK-LLEGLST----- 267
 QY 352 EPDVOTQLDRCOIEPGCGGV--PGAGYDATALVLENOVGNFKOKTLENPDYFNHV 409
 DB 268 --DITPLRLQKESQDQLQVAASSGAGGDCGIALSFDQ---STKTLKN----- 314
 QY 410 YWVDL 414
 DB 315 RWADL 319

RESULT 16

phosphomevalonate kinase [imported] - Streptococcus pneumoniae (strain TIGR4)
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #ext_change 12-May-2003
 C:Accession: E95044
 R:Retzeilin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
 on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,
 neon, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
 A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; M0ID:21357209; PMID:11463916
 A:Accession: E95044
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-335 <KUR>
 A:Cross-references: GB:AE005672; PIDN:AAK74550.1; PID:914971853; GSPDB:GN00164; TIGR:SP4
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: SP0383
 C:Superfamily: mevalonate kinase

Query Match 5.2%; Score 117; DB 2; Length 335;
 Best Local Similarity 22.4%; Pred. No. 1.1;
 Matches 95; Conservative 53; Mismatches 146; Indels 130; Gaps 20;

9 GKAFIAGGYLVLEPIYDAVVTALSSRMHAVITPKGTSIKESRIKISSPOFANGMEYHI 67
 8 GKLWAGGYALIEFGQALIKDIPYMAEIA-----AFSDYRITYSDMFDRA----- 54
 DB 68 SSNTKEPREVOSRINP--PLEATITIVLAYI---OPTFAFLEIIIVSDPGYHSGEDT 120
 DB 55 -----VLRPNPDYSLIOETIALMGDFLAIRGQVLRPFSLIC----- 92
 QY 121 ETKSSNGEKTFLHSRAITTEVEKTKGSSAGLSVVAATSLISHIPVISTNKOILNV 180
 DB 93 -----GKRERGGKFGGLSSGVVAVKALALAY--DVSVDLELPLK 134
 QY 181 AQIAHCYAOCKIGSGFDVAI-YGLIYVRRFOPALINDVFOVLESDEKPEFTEKLE 239
 DB 135 TS-AVLKRGNGSGMDLACIAEDLVLYQSFDR--QKVAAMLEE-----NLATYLE 184
 QY 240 SNWEKHERCTLPYGIKILMGDVKGGSTPKLVSVVLQWKEKEPESSVYVDQLNSANLQ 299
 DB 185 RDW-----GFSI--SQVKPTLECDPLVG---WTRKENVVSHWVOQIKONINQN 227

QY 300 FMKELREMKYSDPEYIKELHDSVEPLTAIKNIRKG-LQALTKSEVP-----IE 352
 Db 228 F-----LTSSKEVTSISVEALEQKSKRIIDQVEVASKLEGS 266
 QY 353 PDVOTOLDRCQEIPEGCVGV--PAGGYDAIAVLLENQVNFQKLTLENDPYHANY 410
 Db 267 TDITPLRLQKESQDLQTVAKSSGAGGDCGIALSPDAQ-----STKTKN-----R 315
 QY 411 WNDL 414
 Db 316 WADL 319

RESULT 17

T07775
 1lipoygenase (EC 1.13.11.12) LX-3 - potato
 C:/Species: Solanum tuberosum (potato)
 C:/Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 08-Oct-1999
 C:/Accession: T07775
 R:/K0101ets, M.V.; Hannapel, D.J.
 Submitted to the EMBL Data Library, June 1996
 A:/Reference number: Z16124
 A:/Accession: T07775
 A:/Status: preliminary; translated from GB/EMBL/DBJ
 A:/Molecule type: mRNA
 A:/Residues: 1-862 <KOL>
 A:/Cross-references: EMBL:U60202; NID:g1407704; PIDN:AB67865.1; PID:g1407705
 A:/Experimental source: cv. Beroлина
 C:/Genetics:
 A:/Gene: LX-3
 C:/Function:
 A:/Description: catalyzes the oxidation of unsaturated fatty acids with a 1,4-cis,cis per
 C:/Superfamily: lipoygenase
 C:/Keywords: fatty acid oxidation; oxidoreductase

Query Match 5.2%; Score 116; DB 2; Length 862;
 Best Local Similarity 18.4%; Pred. No. 5.2;
 Matches 76; Conservative 56; Mismatches 115; Indels 166; Gaps 17;

QY 24 YDAVVTALSRMAVITPKGTSIK-----ESKIKISSPQFA--NG 61
 Db 302 FDSFEDVLKLYEGGIKLPQGPFLKALTDISIPLEILKEIIRTDGKFKPTPVQIQEDKS 361
 QY 62 EMEYHISNTEKREVOARINPFLKATIFIVLVIO--PTEAFDLEIITISDPGYHSGE 118
 Db 362 SMR---TDEEFARMLAGVNP-----VVISRLQEPFKXSQLDSEV-----YGNQV 403
 QY 119 DTETK-----TSNGEKTFLYHRAITEVK 144
 Db 404 STTKHEIENTLDTLIDAIKTNRLYLINHDILMPYVARIWTTNKLKVASRTLLFLQD 463
 QY 145 TGL-----GSAGIVSVVAT-----SLSH 164
 Db 464 DGTMRKVAIELSLPHDGBELGAVSKYVPADQGVSGISWQLAKAVAVANDSGVHQLISH 523
 QY 165 FI-----PNVISTNK--DILHNVAQIAHCYAOKKIGSGFVATAIYGLIYRRPOPAL 215
 Db 524 WINTHAIRPFVATIRKQSLVLFPIKHLHPFRDMM-----INALRQL 570
 QY 216 INDVFOVLESDD--PEKFPTELKLIENSWEEKH-----ERCTLPYGIKILM 259
 Db 571 IN-AGGVLEMTVPFPAKYAMMSAVVYKSWFPEQALPADLIKRGVAVESSSPHVRLLI 629
 QY 260 GDVKGSETPKLVSRVLQWKKK-----KRESSVYVYDQNSANLQPMKELRE 306
 Db 630 QDVPYAVDGLIEIMSAIKSWTEYCNFYXSDLVLDKNDLQIA-----WVKELRE 678

RESULT 18

T32289
 hypotrichal protein F42G2.6 - Caenorhabditis elegans
 C:/Species: Caenorhabditis elegans

C:/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:/Accession: T32289
 R:/Scheet, P.; Magg, L.
 submitted to the EMBL Data Library, September 1997
 A:/Description: The sequence of C. elegans cosmid F42G2.
 A:/Reference number: Z21146
 A:/Accession: T32289
 A:/Status: preliminary; translated from GB/EMBL/DBJ
 A:/Molecule type: DNA
 A:/Residues: 1-833 <SCH>
 A:/Cross-references: EMBL:AF024499; PIDN:AB70358.1; GSPDB:GN00020; CESP:F42G2.6
 A:/Experimental source: strain Bristol N2; clone F42G2
 C:/Genetics:
 A:/Gene: CESP:F42G2.6
 A:/Map position: 2
 A:/Introns: 317/1; 335/1; 362/3; 781/1

Query Match 5.1%; Score 113; DB 2; Length 833;
 Best Local Similarity 27.6%; Pred. No. 8.1;
 Matches 55; Conservative 30; Mismatches 66; Indels 48; Gaps 10;

QY 193 GSGPDVATAYGLIYRRFQPALINDVFOVLESDEKPTTELKLIENSWEEKHERCTL 252
 Db 625 GSGLTISEA-HQTV---GSPMLI-EKKEELGSLHOPVANEIIDLQKE--ETTKP 676
 QY 253 -YGIKLMGDVYGSGSETPKLVSRVLQWKKKEPEES-----SVVYDQ----- 293
 Db 677 EYRKRLIRMSKRRKQNLQKL-----WKEKNTENTHAKTNQEKKLTEMNAVQEDLEQM 730
 QY 294 -----NSANIQFMKELREMKYSDPEYIKELHDSVEPLTAIKNIRKGLQAL 343
 Db 731 KMIQERRRADSQSKNKLQWVLEMDKIDAQ-ETRIKELREVIK-----SDKAKKGEQLT 785
 QY 344 TQKSEVPIEPDQVQQLDR 362
 Db 786 DIRSPNLIKQWQKSAINR 804

RESULT 19

PSKRB
 outer capsid protein VP5 - bluetongue virus (serotype 13, strain USA)
 C:/Species: bluetongue virus
 C:/Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
 C:/Accession: A38481
 R:/Oldfield, S.; Hirasawa, T.; Roy, P.
 J. Gen. Virol. 72, 449-451, 1991
 A:/Title: Sequence conservation of the outer capsid protein, VP5, of bluetongue virus, a
 A:/Reference number: A38481; MUID:91132147; PMID:1847179
 A:/Accession: A38481
 A:/Molecule type: genomic RNA
 A:/Residues: 1-526 <OLD>
 A:/Cross-references: GB:X54308; NID:g60704; PIDN:CAA8205.1; PID:g60705
 C:/Genetics:
 A:/Map position: segment 5
 C:/Superfamily: bluetongue virus outer capsid protein VP5
 C:/Keywords: capsid protein; coat protein

Query Match 5.0%; Score 112.5; DB 1; Length 526;
 Best Local Similarity 21.6%; Pred. No. 4.3;
 Matches 63; Conservative 51; Mismatches 91; Indels 87; Gaps 14;

QY 162 LSHF---IPNVISTN--KDILHNVAQIAHCYAOKKIGS-GPD--VATAIYGLIYRRP-- 211
 Db 8 LSRFGKQGNALTSYAKKIYNTIKAKERPAESRISGAIDGLVQSGVSHIITGESYGE 67
 QY 212 ---QPALIN-----DVFQVLESDEKPTTELKLIENSWEEKHERCTLPG----- 254
 Db 68 SVKQAVLVNLVGAQGEIPDPLSPGGRGIQTIKETIE--EQRNLVIRIKYKGEIRERFG 124
 QY 255 -----IKLMGDVYGSGSETPKLVSRVLQWKKKEPEESSV-----YDQNSANLQPMK 302
 Db 125 EQLEIRIYOFMNGEVVG-----BEQEOYKVLCKAVDSYENLVTENQWR 170

QY 303 EL-----REMKRYDSDPEYTIKELDSVEPLTVAIKIRKGLQALTKQSEVPIDPDVOT 357
 Db 171 TLARALQREATERTEKTESK-MVKEKROKIDALKALIEVERDQOE----- 214
 QY 358 QILDRCOEIPGCGGV-----VPGAGYDAIVLVLENQGNFQKLTEN 402
 Db 215 ---EAIQIAGMTADVLEAASEEVLVSGMATAIATGAIKAGAYLKVKVIN 263

RESULT 20

F72474

probable mevalonate kinase APE2439 - Aeropyrum pernix (strain KI)

C/Species: Aeropyrum pernix

C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999

C/Accession: F72474

R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hatakeyama, Y.; Uchi-no, K.; Takahara, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K

DNA Res. 6, 83-101, 1999

A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum

A/Reference number: A72450; MUID:99310339; PMID:10382966

A/Accession: F72474

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-324 <RAW>

A/Cross-references: DDBJ:AP000064; NID:95105945; PIDN:BA81454.1; PID:dl045240; PID:9510

A/Experimental source: strain KI

C/Genetics:

A/Genes: APE2439

Query Match 5.0%; Score 111.5; DB 2; Length 324;
 Best Local Similarity 22.4%; Pred. No. 2.4;
 Matches 90; Conservative 58; Mismatches 146; Indels 107; Gaps 19;

QY 6 SAPGKFLAGLYLEPIYDAYVYTAISPMHAIITP--KGTISKESRI-KISPGQANEE 62
 Db 8 SAPGKVIIVGEHVVVANGSL-ATVAATGRRLRVTVSSGKIVLESMLGRHSAIPGQG- 65
 QY 63 WEYHISNTEKPREVQSRINPFLIATITVLAIVIOPTFAFDLEIIISDPGYHSQEDTET 122
 Db 66 -----AAKXSPVLEPIY-----AVLRSLARGVSVPH----- 94
 QY 123 KTSNNGEKFVLYHSRAITVEKTKGSSAGLVSVATSLSHPIFNIVSTNKDILHNAQ 182
 Db 95 -----TILVESGI--PPRAGLGSSA--ASMWVA-LSTYSAMHGDLSAEDLYSVAM 140
 QY 183 IAHCAVQKTKIGSGPDVATAIY-GLIYRRFOPALINDVQVLESDEKPEPTLAKLIEN 241
 Db 141 EEEKTAHKG-PSGVDTTIAVRGVLAIRRGEN-----PVDIRPGLT-- 180
 QY 242 WEEKIERCTLPYGIKILMGDVKGSSTPPLVSRVLQWKKEKPESSVYVQDLSANLQPM 301
 Db 181 -----CVTLIVADTGVERTTRVVEHVLISADLGEASTIYI--RAADL-IA 224
 QY 302 KELREKREKYSDDPEYTIKELDSVEPLTVAIKIRKGLQALTKQSEVPIDPDVOT 361
 Db 225 REALHAIEKGD-----ERLGL-IMNAAQGLLSIGASSLEIE-----TLVY 265
 QY 362 RQGEIPGCGVVPAGGAYDAIVLVLENQGNFQKLTEN 402
 Db 266 RRRS-AGALGAKLTAGMGCGVIGLFKESEVERGESVES 305

RESULT 21

H71404

hypothetical protein d13195c - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

A/Variety: Columbia

C/Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000

C/Accession: H71404

R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk

P.; Medler, H.; Medler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terryn, N.; Giel

avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.

Nature 391, 485-488, 1998

A/Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
 etroff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
 C.; Chalatzis, N.
 A/Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
 A/Reference number: A71400; MUID:98121113; PMID:9461215
 A/Accession: H71404
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-1249 <BEV>
 A/Cross-references: GB:Z97335; NID:92244747; PID:92244787

C/Genetics:

A/Map position: 4COP9-4G3845

C/Superfamily: Arabidopsis thaliana hypothetical protein d13195c

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

C/Supernote: Arabidopsis thaliana

A:Reference number: Z22022
 A:Accession: T41942
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-721 <NIC>
 A:Cross-references: EMBL:U43400; PIDN:AAC54702.1
 A:Experimental source: strain J1
 C:Genetics:
 A:Note: U40
 C:Superfamily: herpesvirus infected cell protein ICP18.5

Query Match 5.0%; Score 110.5; DB 2; Length 721;
 Best Local Similarity 20.5%; Pred. No. 9.7;
 Matches 100; Conservative 74; Mismatches 197; Indels 117; Gaps 23;

13 LAGVLYLEP-TYDAYVTALSSRMNAVITP-KGTSIKESRITSSPOFANGMEWHISN 70
 171 LCGVCVMMKEVEFSETTCTLCYEBLSVNOGKSIR---KRLGKFCNHLTETHWVSN 226
 71 TEKPEVOSRINPFEATIFIVLAVI-QTEAFDEIITISDPGYHSGEDTETKTSN-- 127
 227 LKNDVILKODFSTKQGLVKEVMKITTNIPOGQ--LYSKRP-HLGEAENTLINPDLF 283
 128 -----GEKFLYHSRAITP--VEKTGLGSS-----AGLVSVVATSLSHFTPN 168
 284 SKIPPTIYLSSEFTYWSK-ISESVIOKASITLNQNLCHSLYADQNEISKEFLYGETIOD 342
 169 VISTNKDILHNAQAIAHCYAOCKKIGSPVATAIYGLIYRRFPQALINDVFOVLESPP 227
 343 VFNNEE--NTVNDKLY---IGSRP-----ISPCHLVITITVSIKNEEDPV 386
 228 -----EKFPTELKCLIESNMEKHERCTLPYGIKLMGD-----VKGSSETPKL- 271
 367 FTKLAEDEIQRKITLNLNLSNAHEVTPKKVYHSMQDHNLOEIHIRKKAAYOKIS 446
 222 ---VSRVLOMKKEKEBESV-----VDQLSANIQFMKEILREMEKXTDSOPE 316
 447 ESGSKVMKCIKE--QEAALINKLMNINILGNHIFESLSKMNNAFANROQLSLENFADPF 504
 317 TY-----IKELHVSVP-----LTVAIKIRKGLQALTOXSVLEPDPV 356
 505 TYDDHLVYKNNLISKLLPOELLPNLISQEMYRLLTPLSLVHTASFPPLSNISMAVACVA 564
 357 TQLDRCOBIPGCVGVVPGAGYDAIAVLVLENOGVNFKOKTLENPPYFHNVYVWDLLE 416
 565 DFLPMKEDIAKCVGCTI-----YENMMLCTYNNFFNF-----DGLHNIN--DNQR 609
 417 QTEGVLEP 424
 610 QMMNFIRE 617

Db

RESULT 25
 T22849
 hypothetical protein F59F5.7 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T22849; T23019
 R:White, S.
 Submitted to the EMBL Data Library, February 1996
 A:Reference number: Z19625
 A:Accession: T22849
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-790 <W11>
 A:Cross-references: EMBL:Z69646; PIDN:CAA93477.2; GSPDB:GN00028; CESP:F59F5.7
 A:Experimental source: clone F57C7
 R:Lennard, N.
 Submitted to the EMBL Data Library, August 1995
 A:Reference number: Z19653
 A:Accession: T23019
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA

A:Residues: 1-790 <W12>
 A:Cross-references: EMBL:Z50794; PIDN:CAA90661.2; GSPDB:GN00028; CESP:F59F5.7
 A:Experimental source: clone F59F5
 C:Genetics:
 A:Gene: CESP:F59F5.7
 A:Map position: X
 A:Introns: 21/1; 58/3; 101/2; 150/1; 192/3; 270/3; 309/3; 365/1; 411/3; 441/2; 495/3; 60

Query Match 5.0%; Score 110.5; DB 2; Length 790;
 Best Local Similarity 19.7%; Pred. No. 11;
 Matches 101; Conservative 72; Mismatches 179; Indels 161; Gaps 24;

9 GKAFAGGYLYLEPTDAYVTALSSRMNAVIT-----PKGTSIKESR-IKISSPOFAN 60
 199 GVPMTAGWETGQPHYSG-----DRRLHAATFPGSPSIPATVVRKLRKIKNSLFOYN 253
 61 GEMEHISSTNEKREVOSRINPFEATIFIVLAVIOQTEAFD--EITTY----- 109
 254 GE---LASALLQGNVPEPNITSOQATFEKMLANEMSEVKLSKLANMLGFNNGGGTGE 309
 110 SDPGYHSGEDTETKTSNKGKTPLYHSRAITREVEKTGLSGSAGLVSVVATSLSHFTPN 169
 310 SSGTGLFETMTSVLSSGNRK-VPTSSAHDQSQVAPASAGRPQ--TSMLTQFFVKA 365
 170 ISTNKDILHNAQAIAHCYAOCKKIGSPVATAIYGLIYRRFPQALINDVFOVLES--- 225
 366 SDALKD-----SPAKK-----VFENKKNKIID 387
 226 -----DPEK-----FPTELKCLIESNMEKHERCTLPYGIKLMGD 261
 368 PLALPIPEPEKKEKSEVENTLKLSTMPDEORFMESAKIDGIDADSPLIKNLVXND 447
 262 VKGSSETPKLVSRLQWKE-----KPESSVVDQLSANIQFM-KELREMR-KYSD 314
 448 VTESKKEK-ANRLIEMTRSRPSTKTPESGI-----PKAVPYKGLSLAETPIEAK 500
 315 PETYIK-ELDHVSVEPLTVAIKIRKGLQALTOXSEVPIE--PDVQTLDRCOE--IPG 368
 501 PRTVAAVKLFEGIPPIAVNTRVGRDSNVVNAATK-PIELFVEVSTVYQINCMKSNIOG 559
 369 CVGGVPGAGYDAIAVLVLENOGVNFKOKT-----LENPDY----- 405
 560 GIMEVRPENG-----VLKLEKQADTVTSMSSVPGGLRLRQEPDPVDKVPQPLQWPEG 613
 406 -----FHNVYVWDLLEQTEGVLEEK 425
 614 FOPLMLTLPESKVFGDHWSLRDKKNVLDVK 646

Db

RESULT 26
 P89630
 protein F59F5.7 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
 C:Accession: P89630
 R:Anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C.ele
 A:Note: published extrata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A:Accession: P89630
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-829 <STO>
 A:Cross-references: GB:chr_X; PIDN:CAA93477.1; PID:G3877724; GSPDB:GN00028; CESP:F59F5.7
 C:Genetics:
 A:Gene: F59F5.7
 A:Map position: X

Query Match 5.0%; Score 110.5; DB 2; Length 829;
 Best Local Similarity 19.7%; Pred. No. 12;
 Matches 101; Conservative 72; Mismatches 179; Indels 161; Gaps 24;

```

Qy      9 GKAFJAGVLEPIEYDAVYVALSRMAVIT-----PKGTLKSER-IKISSPOAN 60
Db      238 GVPMTAGMWEQOPHYSG-----DRRLHAAPTFGQSPSIPATATVTKRLKIKSNLKOYN 292
Qy      61 GEMEHYSSNTEKPREVOGRINPLEATITVLAVIQTEAFDL--EIIIV----- 109
Db      293 GE---LASALLQOKNVEPNITISQOATPEKMLNEMSVETLSKLANNLLGFNNCGGTGE 348
Qy      110 SDPGVHSEDTEETKSSNGEKTPLYHSRAITREVEKTEGSSAGLVSVAATSLSHPIPNV 169
Db      349 SGSGTGLSEFMTYSTLSGGRNK-VPTSSAHDQSOQYAPASGRPO--TSMLOQFPAKA 404
Qy      170 ISTNKDILHNAQIAHCAQOKKIGSGFDVATAYGLVYRRFPALINDVFOYLES--- 225
Db      405 SDALKD-----SFAKK-----VEENKENKUID 426
Qy      226 -----DPEK-----FTELEKKLIESNMEKHHERCTLYGKILMGD 261
Db      427 PLALPIPEPEKEKEKXKEVENTKLILSTMPEORMTESAIKDEIDADSLIKOLVAND 486
Qy      262 VKGSETPEKVLVSRYLOWMKE-----KPESSVUYDOLNSANLQFM-KELAREME-KYDSD 314
Db      487 VTESKKEK-KANRLIEWIRSNRPSTKTPESGI-----PKNVYPYGKYLGLAETPIEAK 539
Qy      315 PETYIK--ELDHSVEPLTVAIKNIRKGLQALTQKSEVPIE--PDVQQLDRQOE--IPG 368
Db      540 PRTYNAAVKLEPEGIPPIAVNATRVGRDSNVNATAK-PIELFVESTWYQINCKMSNIQG 598
Qy      369 CVGGVVPAGAGYDAIVLVLNVOGNFROKT-----LENPDY----- 405
Db      599 GIMEVRPONG-----VLKLEKQADVTYSMSVPGGLRLQRPBFDVPVDKVPQPLQWMEG 652
Qy      406 -----FANVYVVDLEBQTEGVLEEK 425
Db      653 FQPLMLTLPSKQVFGDHIWVSLRDHKRNILPVK 685

```

RESULT 27
C71106
hypothetical protein PH0620 - *Pyrococcus horikoshii*
C:Species: *Pyrococcus horikoshii*
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
C:Accession: C71106
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekita
M.; Ohkuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:96344137; PMID:9679194
A:Accession: C71106
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-664 <KAW>
A:Cross-references: GB:AP000003; NID:g3236130; PIDN:BAA29709.1; PID:d1030652; PID:g3257570
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
;Gene: PH0620

Query Match 4.9%; Score 110; DB 2; Length 664;
Best Local Similarity 20.6%; Pred. No. 9.3;
Matches 73; Conservative 61; Mismatches 131; Indels 90; Gaps 18

```

Qy 34 MHAHATTPGTSJKS-----RIKISPOCANMEWEHISNTEPREVOSRINPFL 85
    ||::||::||::||::||::||::||::||::||::||::||::||::||
Db 156 RHSHIQAQVREIEELRRADIPFDEIEERDVGSRGERRYVAGB--ELAGLVKPMR 212
    ||::||::||::||::||::||::||::||::||::||::||::||
Qy 86 EATIFIVLAYIOPTAEFDE-----IIISDPGYHSQEDTETKSSNGEK 130
    |||||::|||::|||::|||::|||::|||::|||::|||::|||
Db 213 GGDVET--RIQPVKRVLEFVPLKETAIEERKSLIVGLDPPETVGLAIDLDGNIIS 268
    |||||::|||::|||::|||::|||::|||::|||::|||::|||
Qy 131 TFLYHSRAITEVEK-----TGLGSSAGLVSVVATSLISH-FIEN--VISTNK 174
    ||::||::||::||::||::||::||::||::||::||::||::||

```

```

D6 269 VSEKMMALSELVIRFISELGHFIYATVDNPAFGVEKNAISFPAQOLFVPRESUKAVEKN 328
QY 175 DILHNAQAIAHCYAOKKIGSGFVATAIYGLIVRRFOPALINDVFOVLESDEKPFTEL 234
D6 329 ELKKNIGISVSDHQ-----DALAAAYK-AYLRYKPKL-----EHIEARL 368
QY 235 KKLIESNBEKHRCRLPYGILKLMGDVKGSGEFPKLVSRLOKKEKPESSVIVYQOLN 298
D6 369 KEL-GIMKKRNE-----IKAL-----ILSGSILGEALIMKYK--TOEKPRE-EVITEREB 413
QY 295 SAULO-FMKELNEMEKYDSDPEYIKELJDSVEPLTAIINIRKGOALTOKSE 348
D6 414 SVDIKFYLKRIEEL-----ERVYBELERNSLKAILEOKKILIEKLERKJE 460

```

RESULT 28
B75055
hypothetical protein PAB1429 - *Pyrococcus abyssi* (strain Orsay)
C:Species: *Pyrococcus abyssi*
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: B75055
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome structure
A:Reference number: A75001
A:Accession: B75055
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-643 <KW>
A:Cross-references: GB:AJ248287; GB:AL096836; NID:95458657; PIDN:CBAS0335.1; PID:el516233
A:Experimental source: strain Orsay
C:Genetics:
C:Gene: PAB1429

Query Match	4.9%	Score 109.5	DB 2	Length 643
Best Local Similarity	19.9%	Pred. No. 9.5		
Matches 73	Conservative 64	Mismatches 140	Indels 89	Gaps 17

```

Qy 34 RMAVATPPTGTSLKES-----RIKISSQFANGWEVHISNTKEPRVQSRINPFL 85
Db 136 RVHSLLOAKVROQIEELKRAPIPEPDLIEBERDVYSGRGEFVVASR---ELAGLVPMR 192
Qy 86 EAFIFIVAVIOTFEAFDLE-----IIISDPGHSQEDETTKTSNGEK 130G
Db 193 GGVDEL---RIQVBERKVLEFVPLKAERKERSIIVGLDPGI-TVGIAIDIDGAVIA 248G
Qy 131 TELYHSRAITEYEK-----TGLGSAGIIVSVATSLSH-PIPN--VISTNK 174G
Db 249 VYSEKMMALSEIYRFISELGHPIIVATVPSAPAGIYVERKIASFRAOLFVPRSLKVEBK 308G
Qy 175 DILAHVQAOLAHCAOKKIGSGFDVATATYGLIYVRPOPALINDVQFLSDPEKPFTEL 234G
Db 309 ELIRNIGISVSDHQR-----DALAAVK-AYLRVYPKL-----EHEVAKL 348G
Qy 235 KKLIESNWEKHERCTLPYGIKILMGDVYGSSETPKLSRVLOWKKEKPEBSSVVDOLN 294G
Db 349 KEY-GLWKKRND-----IKAL---ILAGSISGEALMKIK-AKEKPREKTERREE 394G
Qy 295 SANIOWMKELREMERKEDYSDPEPTYIKELDHSVETPTVAIKIRKGLQALQKSEVPIEPD 354G
Db 395 VDVEQYIRRIEEL-----EKTVKELEREVAELRAIIEORKIIEKLEBKELE-BFDEN 444G
Qy 355 VOTQOL 360
Db 446 VRUKVL 451

```

RESULT 29
T52176
aldelyde oxidase (EC 1.2.3.1) [imported] - Arabidopsis thaliana (firemint)
C|Species: Arabidopsis thaliana (mouse-ear cress)
C|Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C|Accession: T52176

R.Sekimoto, H.; Seo, M.; Kawakami, N.; Komano, T.; Desloire, S.; Liotenberg, S.; Marion-Plant Cell Physiol. 39, 433-442, 1998
 A:Title: Molecular cloning and characterization of aldehyde oxidases in Arabidopsis thaliana
 A:Reference number: 225420, MUID:98277700, PMID:9615466
 A:Accession: T52176

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-986 <SEK>

A:Cross-references: EMBL:AB010080; PIDD:BA28630.1
 C:Genetics:

A:Gene: atAO-3
 C:Keywords: oxidoreductase

Query Match 4.9%; Score 109.5; DB 2; Length 986;
 Best Local Similarity 20.0%; Pred. No. 18;
 Matches 99; Conservative 69; Mismatches 179; Indels 147; Gaps 23;

```

QY 4 AASAPKAF-----LAGGYLV-LEPIYDAYVTLSSR-----MAAVTPK-- 42
DB 229 ALQASGEAVFVDIPPLPDLCHGAFYSTEP--AKIKSLSFRENVTPGVAVLTFKDI 286
QY 43 ---GTSLSKSRKISSPOFAN-----GEMEHISNTEKPREVOSRI-----NPFLFA 87
DB 287 PQGGNIGSKTLPFGPPLFADBLTRCAGQRIALVADTQKADMAKLAVEREYDKNLEQ 346
QY 88 TFIYLVAYIOPTFAFDLEIIYSDP---GYHSQEDTETKSSN---GEKTFLYHSRAIT 140
DB 347 PLTVEDAVKSSPFVHMFPEPVGDVYKGMEEAEKRIISSELRIGSQYFFY---ME 402
QY 141 EYEKIGSSAGLVSVATSLSHFIPNVISTNKDIL-HNVAQLAHCAQKIGSGF--- 196
DB 403 PQTALALPDEDNCVAVFSSQAPREVHSAVATCLGIOEHNVAVIT---RRVGGGFGGK 457
QY 197 -----DVATATYGLIVYRRFOPALI-----NDVFQV-----LESDEKPFPE 233
DB 458 AVKSNPVAATAC-ALGAYVQLQRPVKMFLNKRKTMIMAGGRHMKINYNVGFSDGLTALF 516
QY 234 LKKLIESNWE-----EKHERCTLPYGIKLLM-----GDVKGGS 266
DB 517 LTMILDAGLEPDVSPIMPRNIMGPLARKYDGMALSPDVKCKTNCLSRAMRPAgevQSY 576
QY 267 ETPKLVSRVLQWKKEKPESSVVDQLNSANLQFMKELEKREKTDSPETV----- 318
DB 577 IAEIIEINV-----ASSIQMDVDAVRKINLHTYDSLARKFYNHLAGDPDEYTLPLMEK 629
QY 319 -----IKELDSVEPLTVAIKIRKGLQALTKQSEVPIEBPVQQLDRCEIIGCV-- 370
DB 630 LEISSKFKERSSEWKEFNLGNVWRKRG-----SRPVIHQVMO-----RPTGKXSI 677
QY 371 ---GGVVPAGGYD 381
DB 678 LSDGSVVVEVGIE 691

```

RESULT 30

adenyde oxidase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 23-Mar-2001

C:Accession: D84669

R.Liim, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Umayam, L.; Tallon, L.; Neus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: AB4420; MUID:20083487; PMID:10617197

A:Accession: D84669

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1332 <STO>

A:Cross-references: GB:AB002093; NID:g4557058; PIDD:AA022498.1; GSPDB:GM00139

C:Genetics:

A:Gene: At2g27150

A:Map position: 2
 C:Superfamily: xanthine dehydrogenase; ferredoxin [2Fe-2S] homology
 C:Keywords: 2Fe-2S; metalloprotein
 F:40,45,48,70/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 4.9%; Score 109.5; DB 2; Length 1332;
 Best Local Similarity 20.0%; Pred. No. 29;
 Matches 99; Conservative 69; Mismatches 179; Indels 147; Gaps 23;

```

QY 4 AASAPKAF-----LAGGYLV-LEPIYDAYVTLSSR-----MAAVTPK-- 42
DB 575 ALQASGEAVFVDIPPLPDLCHGAFYSTEP--AKIKSLSFRENVTPGVAVLTFKDI 632
QY 43 ---GTSLSKSRKISSPOFAN-----GEMEHISNTEKPREVOSRI-----NPFLFA 87
DB 633 PQGGNIGSKTLPFGPPLFADBLTRCAGQRIALVADTQKADMAKLAVEREYDKNLEQ 692
QY 88 TFIYLVAYIOPTFAFDLEIIYSDP---GYHSQEDTETKSSN---GEKTFLYHSRAIT 140
DB 693 PLTVEDAVKSSPFVHMFPEPVGDVYKGMEEAEKRIISSELRIGSQYFFY---ME 748
QY 141 EYEKIGSSAGLVSVATSLSHFIPNVISTNKDIL-HNVAQLAHCAQKIGSGF--- 196
DB 749 PQTALALPDEDNCVAVFSSQAPREVHSAVATCLGIOEHNVAVIT---RRVGGGFGGK 803
QY 197 -----DVATATYGLIVYRRFOPALI-----NDVFQV-----LESDEKPFPE 233
DB 804 AVKSNPVAATAC-ALGAYVQLQRPVKMFLNKRKTMIMAGGRHMKINYNVGFSDGLTALF 862
QY 234 LKKLIESNWE-----EKHERCTLPYGIKLLM-----GDVKGGS 266
DB 863 LTMILDAGLEPDVSPIMPRNIMGPLARKYDGMALSPDVKCKTNCLSRAMRPAgevQSY 922
QY 267 ETPKLVSRVLQWKKEKPESSVVDQLNSANLQFMKELEKREKTDSPETV----- 318
DB 923 IAEIIEINV-----ASSIQMDVDAVRKINLHTYDSLARKFYNHLAGDPDEYTLPLMEK 975
QY 319 -----IKELDSVEPLTVAIKIRKGLQALTKQSEVPIEBPVQQLDRCEIIGCV-- 370
DB 976 LEISSKFKERSSEWKEFNLGNVWRKRG-----SRPVIHQVMO-----RPTGKXSI 1023
QY 371 ---GGVVPAGGYD 381
DB 1024 LSDGSVVVEVGIE 1037

```

RESULT 31

apolipoprotein B - pig (fragments)

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 13-Aug-1999

C:Accession: J70382; I46567; I46568

R.Maeda, N.; Ebert, D.L.; Doers, T.M.; Newman, M.; Hasler-Rapacz, J.; Attie, A.D.; Rapacz, Gene 70, 213-229, 1988

A:Title: Molecular genetics of the apolipoprotein B gene in pigs in relation to atherosclerosis

A:Reference number: J70382; MUID:89108006; PMID:2905687

A:Accession: J70382

A:Molecule type: DNA

A:Residues: 1-1778 <MA>

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-8, 'S', 10-238 <MA>

A:Cross-references: GB:M22646; NID:g164366; PIDD:AAA30996.1; PID:g164369

A:Accession: I46568

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 239-1778 <MA>

A:Cross-references: GB:M22647; NID:g164367; PIDD:AAA30997.1; PID:g164370

C/Comment: Apolipoprotein B is the predominant protein component of the low-density lipoprotein
 C/Genetics:
 A:Gene: apob
 A:Introns: 39/3; 88/3; 159/2; 238/3; 1133/1; 1171/2; 1232/3
 A/Note: the list of introns may be incomplete
 C/Superfamily: apolipoprotein B
 C/Keywords: atherosclerosis; cholesterol metabolism; LDL; lipid binding; lipoprotein

Query Match 4.9%; Score 108.5; DB 2; Length 1778;
 Best Local Similarity 20.7%; Pred. No. 53;
 Matches 69; Conservative 53; Mismatches 122; Indels 101; Gaps 16;

QY 41 PKGTSLE-----SRKISSPPFANGEMEHISNTKEPREVQSRINPFEATIFVLAY 95
 DB 1075 PAGESLPSFGTLARAGVASPLV-NATWSAGLKNK-----EDRETFLDSCSTIQF 1126
 QY 96 IQPTEAFDEIIITSPGCHSOEDTETKSSNGEKTFLYHSRAITVEKT---GLGSSA 151
 DB 1127 LE---YDINLV---GTHKIEDGMLVCSITKGS---FAHRDLAAYEENGKHYGLEWYD 1174
 QY 152 GLVSVVATSLSHFIPNVISTNMDILHNAQIHCAQKKIGSGFDVATAYGLIYVRB- 210
 DB 1175 GKIIITLTS-----PLTIDVYLR-----YQSKNMLFSSVASPSIGTVALBOE 1217
 QY 211 -POPALINDVFOVLESDEPK-----PTELEKLT-----IESNWEKHERCTLPYGIKLT 258
 DB 1218 NINTSLKSNLYRPPQSPDKNIFIKERKLDSESDVOIKANWEE----- 1263
 QY 259 MGDVKGGSETPKLVSVLQMKKEK-PESSVVTYDQIN-----SANTQFMKEIR 305
 DB 1264 -----EAVSELSRLSKDNVPEKATGAFDYVNAKHOEYTGDLRLDASIKLRKQ 1312
 QY 306 EMREKYSDEPTIKELDHSVERPL-----TVAIKNIKIGQALQLOK 346
 DB 1313 NSADRVHQAARRIDEDVRLKRIKATGTTKYQOMKDRQSLYQK 1357

RESULT 32

S28261
 centromere protein E - human
 N/Alternate names: centromere 312K protein; kinesin-related protein CENP-E
 C/Species: Homo sapiens (man)
 C/Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
 C/Accession: S28261
 R/Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
 Nature 359, 536-539, 1992
 A/Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
 A/Reference number: S28261; MUID:93024922; PMID:1406971
 A/Accession: S28261
 A/Molecule type: mRNA
 A/Residues: 1-2663 <YEN>
 A/Cross-references: EMBL:Z15005; NID:g29864; PIDD:CAA78727.1; PID:g29865
 C/Genetics:
 A:Gene: GDB:CENPE
 A/Cross-references: GDB:361164; OMIM:117143
 A/Map position: 4q24-4q25
 C/Superfamily: centromere protein E; kinesin motor domain homology
 C/Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
 F/7-35/Domain: kinesin motor domain homology <KNOT>
 F/86-93/Region: nucleotide-binding motif A (P-loop)
 F/486-2183/Domain: coiled coil #status predicted <COI>
 F/92/Binding site: ATP (Lys) #status predicted

Query Match 4.9%; Score 108.5; DB 1; Length 2663;
 Best Local Similarity 20.7%; Pred. No. 98;
 Matches 84; Conservative 67; Mismatches 161; Indels 93; Gaps 16;

QY 47 KSRRIKISSPPFANGEMEHISNTKEPREVQSRINPFEATIFVLAYIQPEARDLEI 106
 DB 852 KEAQKDDSLGALKTELSTKTOGLKTRERVDRLAE-MEQLKEQLENDSPLOQVREK 910
 QY 107 IYSDPGVHSDDETEKTSNGEKTFLYHSRAITVEKTLGSSA----- 151

DB 911 TLTEKLOQTLEEVKTLTQ---EKDQLQLOESIQIERDQLSDIHDYVNNINDTQEOQLR 967
 QY 152 -GLVSV-----VATSLSHFIPNVISTNMDILHNAQIHCAQKKIG-----SGFD 197
 DB 968 NALSELKHOETINTLKRSISEVSRNLMHEENTETDEREQKRVGIDKQDLAKTQ 1027
 QY 198 VATV-----IYGLIYRRPOPALINDVFOVLES---DPEKPELEKTLISN 241
 DB 1028 TLTADVKNNEIEQQRKIFSLIQEK-----NELQQLVESVIAKEKQDLKNDKENT 1080
 QY 242 WEEKHERCTLPYGIKLMGVDVKGSETPLVSRVLQMKK---EKPESSVVTYDQINSANL 298
 DB 1081 IENQEE-----LRLT-----GDELKQOEIVAQEKNHAIKKEGELSRCTDRLAEEYR 1127
 QY 299 QFMKELREMERKYS--DPEYIKELDHSVERPLVVAIKNIKIGQALQLOKSEVPTEPDVQ 356
 DB 1128 KKEKESQQLQOEQOQLLVQESMSQKKIN---EIEELKNEL---KNEELTLE-HWE 1178
 QY 357 TQLDRCOEIPQCVGAVVPGAGYDAIYVLVLEQVGNFKOKTLE 401
 DB 1179 TERLELAQKL-----NENYEVKSTITERKVLKELQKSF 1213

RESULT 33

B70419
 hypothetical protein aq_1369 - Aquifex aeolicus
 C/Species: Aquifex aeolicus
 C/Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 04-Mar-2000
 C/Accession: B70419
 R/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
 V.
 Nature 392, 353-358, 1998
 A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A/Reference number: A70300; MUID:98196666; PMID:9537320
 A/Accession: B70419
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-260 <AOF>
 A/Cross-references: GB:AE000737; NID:g2983782; PIDD:AAO07356.1; PID:g2983798; GB:AE00065
 A/Experimental source: strain VFS
 C/Genetics:
 A:Gene: aq_1369
 C/Superfamily: Aquifex aeolicus hypothetical protein aq_1369

Query Match 4.8%; Score 108; DB 2; Length 260;
 Best Local Similarity 24.8%; Pred. No. 3.1; Indels 82; Gaps 13;
 Matches 58; Conservative 26; Mismatches 68; Indels 82; Gaps 13;

QY 205 LLYVR-----RPOPALINDVFOVLESDEPKFTELKQLIESNWEKHERCTLPYGIKLMG 260
 DB 85 LVTVRIPEKEPEKEL--GIFQVE---EAMGAFSLTALHGWEE-----VPKNVVIYHA 133
 QY 261 D-VKGGSETPKLVSVLQMKKEKPESSVVTYDQINSANLQFMKEIREMERK----- 310
 DB 134 DVEBGN--KLIAAI-----KTEBGISTYDQI-----KLEEMKQKVRPRVV 175
 QY 311 YSDDEYIKELDHSVERPLVVAIKNIKIGQALQLOKSEVPTEPDVQTLDRCOEIPGV 370
 DB 176 YSSDVLYIKD-----IYPDQSAAYIYARIKAE 206
 QY 371 GGV--VPGAGYDAIYVLVLEQVGNFKOKTLEND-----DYFNAYVWDL 415
 DB 207 GSAPLEBELGKTYGVDSTLSEKL-ELLEKLLQNVKLPGGKEVNLKRYWYPL 259

RESULT 34

G85564
 phenylalanyl tRNA synthetase beta [imported] - Chlamydia pneumoniae (strain J138)
 C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
 C/Accession: G85564
 R/Shital, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
 Nucleic Acids Res. 28, 2311-2314, 2000

A>Title: Comparison of whole genome sequences of chlamydia pneumoniae UJ38.
A:Reference number: AB6491; MUID:20330349; PMID:10871362
A:Accession: G86564
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-792 <STD>
A:Cross-references: GB:A000008; NID:g8978966; PIDN:BAA98801.1; GSPDB:GN00142
A:Experimental source: strain UJ38
C:Genetics:
A:Gene: phet
C:Superfamily: phenylalanine-tRNA ligase beta chain

Query Match 4.8%; Score 108; DB 2; Length 792;
Best Local Similarity 19.7%; Pred. No. 17;
Matches 75; Conservative 69; Mismatches 133; Indels 104; Gaps 18;

OY 36 HAVITPKGTSLKSRKIKISSPOFANGEMWEHISNTKEPREVOGRINPFLEATTFIYLAY 95
 ||||| : | : | : || : : : : || : : : | : | :
Db 221 YVVT -GISAQPSPIKL-----QESLQALKKP-----INAIYDITYIMLSL 262

OY 96 IQPFAFLFIITISDPCHSQEDTETKTSSNGEKFTLYHSRAITEVEKT--GLGSAG- 152
 ||||| : | : | : || : ||||| : ||||| : ||||| : ||||| :
Db 263 GQPHAYASHVAALDSIVREKISTPESELTLLNGEFTVLIPSGVPVRDDSHSLGLGWAGA 322

OY 153 ----LVSVVAASLL-SHEIPNVISTNKDILAHNAQAHCQAOKKIGSGPVPAIAIIGLI 206
 ||||| : | : | : || : ||||| : ||||| : ||||| : ||||| :
Db 323 KAPSFQETTTTVIKAALFPALPASQKLPPIBESAYRFR-----GIDPQNV----- 373

OY 207 VYRRPOPALINDVQVLESDEKF-----PTELKKLIESNWEEKH 246
 ||||| : | : | : || : ||||| : ||||| : ||||| : ||||| :
Db 374 -----PALQAAIHYLEIFPEATISPIYSSGEICRELKAEVALPKTLQRLLGKS----- 422

OY 247 ERCTLPYGKILMGDVK--GGSETPK--LVSRYLQWKEKEPDESSVVYD--QLNSANLQ 299
 ||||| : | : | : || : ||||| : ||||| : ||||| : ||||| :
Db 423 -----FSIEILISQKLGIGSTPQETSLLVKPVSYNHDINEIDLVEICRTESWNIE 476

OY 300 FMKELEREREKYDDPER-----YIKELDHSVLEPLTAIKIRKGLQALTKS 347
 ||||| : | : | : || : ||||| : ||||| : ||||| : ||||| :
Db 477 TONPSCYTPIYKLKRRTAFGLANANGLOEFFTPDL---LDDETVAL---TRK-----EKE 525

OY 348 EVPIEPDVQTOLDRCQRIPG 368
 || : : : | : | : || : ||||| : ||||| : ||||| : ||||| :
Db 526 EISLQSKHTIVL-RSSLIPG 545

RESULT 35
D81608
phenylalanine-tRNA ligase [EC 6.1.1.20] beta chain CP0154 [similarity] - Chlamydophila P
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: D81608
R:Read, T.D.; Brumham, R.C.; Shen, C.; Gall, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A>Title: Genome sequence of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: D81608
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-792 <REA>
A:Cross-references: GB:A0002176; GB:A0002161; NID:g7189080; PIDN:AAF38035.1; PID:g7189080
A:Experimental source: strain AK39, HL cells
C:Genetics:
A:Gene: CP0154
C:Superfamily: phenylalanine-tRNA ligase beta chain
C:Keywords: aminocyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 4.8%; Score 108; DB 2; Length 792;
Best Local Similarity 19.7%; Pred. No. 17;
Matches 75; Conservative 69; Mismatches 133; Indels 104; Gaps 18;

OY 36 HAVITPKGTSLKSRKIKISSPOFANGEMWEHISNTKEPREVOGRINPFLEATTFIYLAY 95
 ||||| : | : | : || : : : : || : : : | : | :
Db 221 YVVT -GISAQPSPIKL-----QESLQALKKP-----INAIYDITYIMLSL 262

Db 221 YVVI--GISAQSPRIKL-----QESLQALQKP-----INAVDITVIMLSL 262

Qy 96 IQPTAFPLEIIIVSDPGYHSGEDTETKTSSNGEKTFLYHSRAITEVEKT--GLGSSAG- 152

Db 263 GQPLHAYASHVALDSLRLVEKLTSTPESLTLLNGETVLLPFGVAPVRDHSILGLGCVWGA 322

Qy 153 ----LVSVAATSL--SHFIPNVISTNNKDILHVAQIAHCYAAQKIGSGSPVATAIYGLI 206

Db 323 KAPSFQETTTTIVIAAAYFLPEALRASQKLLPISSSAVRFTR-----GIDPQNVV----- 373

Qy 207 VYRRFQPALINDVFOVLESDEPEK-----PTELAKLIESNMWEKH 246

Db 374 -----PALQAAIHVILIEIFPEATISPIYSSGEICRELKVALRPKTLQRLIGKS----- 422

Qy 247 ERCTLPYGIKILMGDVK--GGSSETPK--LVSRVLQWKKKEPESSVYVD--QLNSANLQ 299

Db 423 -----FIEILSQKIQSLGFSFTTPOETSLVLVPSVRHDINEEIDIVEICRTESWNIE 476

Qy 300 FMKELRERREKVDSDPER-----YIKELHSAVEPLVAIKNIKRGALQOKS 347

Db 477 TQNPVSCYTPYIKLRRTAGFLANGLQGFPTPDL--LDPEVAL--TRK-----EKE 525

Qy 348 EVPIBPDVQOTLLDRCOEIRG 368

Db 526 EISLQSKHTIVL-RSSLLPG 545

RESULT 36

H72058

phenylalanine-tRNA ligase (EC 6.1.1.20) beta chain - Chlamydothila pneumoniae (strain CW

C/Species: Chlamydothila pneumoniae, Chlamydia pneumoniae

C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #ext_change 03-Jun-2002

C/Accession: H72058

R/Katman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;

Nature Genet. 21, 385-389, 1999

A/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A/Reference number: A72000; MUID:99206606; PMID:10192388

A/Accession: H72058

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-792 <ARN>

A/Cross-references: GB:AE001644; GB:AE001363; NID:g4376888; PID:AD18733.1; PID:g4376888

A/Experimental source: strain CML029

C/Genetics:

C/Name: pher

C/Superfamily: phenylalanine-tRNA ligase beta chain

C/Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 4.8%; Score 108; DB 2; Length 792;

Best Local Similarity 19.7%; Pred. No. 17;

Matches 75; Conservative 69; Mismatches 133; Indels 104; Gaps 18;

Qy 36 HAVITPKTSLKESRIKISSPOFANGMEYHISNTEKREVOGRINPFEATIFVILAY 95

Db 221 YVVI--GISAQSPRIKL-----QESLQALQKP-----INAVDITVIMLSL 262

Qy 96 IQPTAFPLEIIIVSDPGYHSGEDTETKTSSNGEKTFLYHSRAITEVEKT--GLGSSAG- 152

Db 263 GQPLHAYASHVALDSLRLVEKLTSTPESLTLLNGETVLLPFGVAPVRDHSILGLGCVWGA 322

Qy 153 ----LVSVAATSL--SHFIPNVISTNNKDILHVAQIAHCYAAQKIGSGFVATAIYGLI 206

Db 323 KAPSFQETTTTIVIAAAYFLPEALRASQKLLPISSSAVRFTR-----GIDPQNVV----- 373

Qy 207 VYRRFQPALINDVFOVLESDEPEK-----PTELAKLIESNMWEKH 246

Db 374 -----PALQAAIHVILIEIFPEATISPIYSSGEICRELKVALRPKTLQRLIGKS----- 422

Qy 247 ERCTLPYGIKILMGDVK--GGSSETPK--LVSRVLQWKKKEPESSVYVD--QLNSANLQ 299

Db 423 -----FIEILSQKIQSLGFSFTTPOETSLVLVPSVRHDINEEIDIVEICRTESWNIE 476

Qy 300 FMKELRERREKVDSDPER-----YIKELHSAVEPLVAIKNIKRGALQOKS 347

Db 477 TQNPVSCYTPYIKLRRTAGFLANGLQGFPTPDL--LDPEVAL--TRK-----EKE 525

Qy 348 EVPIBPDVQOTLLDRCOEIRG 368

Db 526 EISLQSKHTIVL-RSSLLPG 545

RESULT 36

H72058

phenylalanine-tRNA ligase (EC 6.1.1.20) beta chain - Chlamydothila pneumoniae (strain CW

C/Species: Chlamydothila pneumoniae, Chlamydia pneumoniae

C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #ext_change 03-Jun-2002

C/Accession: H72058

R/Katman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;

Nature Genet. 21, 385-389, 1999

A/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A/Reference number: A72000; MUID:99206606; PMID:10192388

A/Accession: H72058

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-792 <ARN>

A/Cross-references: GB:AE001644; GB:AE001363; NID:g4376888; PID:AD18733.1; PID:g4376888

A/Experimental source: strain CML029

C/Genetics:

C/Name: pher

C/Superfamily: phenylalanine-tRNA ligase beta chain

C/Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 4.8%; Score 108; DB 2; Length 792;

Best Local Similarity 19.7%; Pred. No. 17;

Matches 75; Conservative 69; Mismatches 133; Indels 104; Gaps 18;

Qy 36 HAVITPKTSLKESRIKISSPOFANGMEYHISNTEKREVOGRINPFEATIFVILAY 95

Db 221 YVVI--GISAQSPRIKL-----QESLQALQKP-----INAVDITVIMLSL 262

Qy 96 IQPTAFPLEIIIVSDPGYHSGEDTETKTSSNGEKTFLYHSRAITEVEKT--GLGSSAG- 152

Db 263 GQPLHAYASHVALDSLRLVEKLTSTPESLTLLNGETVLLPFGVAPVRDHSILGLGCVWGA 322

Qy 153 ----LVSVAATSL--SHFIPNVISTNNKDILHVAQIAHCYAAQKIGSGFVATAIYGLI 206

Db 323 KAPSFQETTTTIVIAAAYFLPEALRASQKLLPISSSAVRFTR-----GIDPQNVV----- 373

Qy 207 VYRRFQPALINDVFOVLESDEPEK-----PTELAKLIESNMWEKH 246

Db 374 -----PALQAAIHVILIEIFPEATISPIYSSGEICRELKVALRPKTLQRLIGKS----- 422

Qy 247 ERCTLPYGIKILMGDVK--GGSSETPK--LVSRVLQWKKKEPESSVYVD--QLNSANLQ 299

Db 423 -----FIEILSQKIQSLGFSFTTPOETSLVLVPSVRHDINEEIDIVEICRTESWNIE 476

Qy 300 FMKELRERREKVDSDPER-----YIKELHSAVEPLVAIKNIKRGALQOKS 347

Db 477 TQNPVSCYTPYIKLRRTAGFLANGLQGFPTPDL--LDPEVAL--TRK-----EKE 525

Qy 348 EVPIBPDVQOTLLDRCOEIRG 368

Db 526 EISLQSKHTIVL-RSSLLPG 545

RESULT 36

H72058

phenylalanine-tRNA ligase (EC 6.1.1.20) beta chain - Chlamydothila pneumoniae (strain CW

C/Species: Chlamydothila pneumoniae, Chlamydia pneumoniae

C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #ext_change 03-Jun-2002

C/Accession: H72058

R/Katman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;

Nature Genet. 21, 385-389, 1999

A/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A/Reference number: A72000; MUID:99206606; PMID:10192388

A/Accession: H72058

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-792 <ARN>

A/Cross-references: GB:AE001644; GB:AE001363; NID:g4376888; PID:AD18733.1; PID:g4376888

A/Experimental source: strain CML029

C/Genetics:

C/Name: pher

C/Superfamily: phenylalanine-tRNA ligase beta chain

C/Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 4.8%; Score 108; DB 2; Length 792;

Best Local Similarity 19.7%; Pred. No. 17;

Matches 75; Conservative 69; Mismatches 133; Indels 104; Gaps 18;

Qy 36 HAVITPKTSLKESRIKISSPOFANGMEYHISNTEKREVOGRINPFEATIFVILAY 95

Db 221 YVVI--GISAQSPRIKL-----QESLQALQKP-----INAVDITVIMLSL 262

Qy 96 IQPTAFPLEIIIVSDPGYHSGEDTETKTSSNGEKTFLYHSRAITEVEKT--GLGSSAG- 152

Db 263 GQPLHAYASHVALDSLRLVEKLTSTPESLTLLNGETVLLPFGVAPVRDHSILGLGCVWGA 322

Qy 153 ----LVSVAATSL--SHFIPNVISTNNKDILHVAQIAHCYAAQKIGSGFVATAIYGLI 206

Db 323 KAPSFQETTTTIVIAAAYFLPEALRASQKLLPISSSAVRFTR-----GIDPQNVV----- 373

Qy 207 VYRRFQPALINDVFOVLESDEPEK-----PTELAKLIESNMWEKH 246

Db 374 -----PALQAAIHVILIEIFPEATISPIYSSGEICRELKVALRPKTLQRLIGKS----- 422

Qy 247 ERCTLPYGIKILMGDVK--GGSSETPK--LVSRVLQWKKKEPESSVYVD--QLNSANLQ 299

Db 423 -----FIEILSQKIQSLGFSFTTPOETSLVLVPSVRHDINEEIDIVEICRTESWNIE 476

Qy 300 FMKELRERREKVDSDPER-----YIKELHSAVEPLVAIKNIKRGALQOKS 347

Db 477 TQNPVSCYTPYIKLRRTAGFLANGLQGFPTPDL--LDPEVAL--TRK-----EKE 525

Qy 348 EVPIBPDVQOTLLDRCOEIRG 368

Db 526 EISLQSKHTIVL-RSSLLPG 545

RESULT 36

H72058

phenylalanine-tRNA ligase (EC 6.1.1.20) beta chain - Chlamydothila pneumoniae (strain CW

C/Species: Chlamydothila pneumoniae, Chlamydia pneumoniae

C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #ext_change 03-Jun-2002

C/Accession: H72058

R/Katman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;

Nature Genet. 21, 385-3

Db 477 TQWVSCYPIYKLTKEKTAGFLNAGLQSEFTPDL---LDPEVTAL--TRK-----EKE 525
 QY 348 EVDIEPDVQTLDRQCEIRPG 368
 Db 526 EISLQSSKHTTVL-RSSLFPG 545

RESULT 37

A29232
 101K malaria antigen precursor - malaria parasite (Plasmodium falciparum) (strain Camp)
 C/Species: Plasmodium falciparum
 C/Date: 22-Dec-1988 #sequence_revision 22-Dec-1988 #ext_change 09-Jun-2000
 C/Accession: A29232
 R/Reber, J.L.; Lyon, J.A.; Wolff, R.H.; Hall, T.; Lowell, G.H.; Chulay, J.D.
 J. Biol. Chem. 263, 11421-11425, 1988
 A/Title: Primary structure of a Plasmodium falciparum malaria antigen located at the mzx
 A/Reference number: A29232; MUID:88298794; PMID:3042768
 A/Accession: A29232
 A/Status: preliminary; not compared with conceptual translation
 A/Molecule type: DNA
 A/Residues: 1-743 <MEB>
 A/Cross-references: GB:J03902; NID:G160046; PID:G160047

Query Match 4.8%; Score 107.5; DB 2; Length 743;
 Best Local Similarity 21.0%; Pred. No. 16;
 Matches 82; Conservative 50; Mismatches 117; Indels 141; Gaps 17;

QY 126 SNGEFTLYHSRAITE---VEKTGLSSAGLVSVATSL-----LSHFIEN 168
 Db 322 SMDLINFLOQKEIGENLNTTQNLGDKNNLESPLDEINMLKDNINVEFLDMLKTS 381
 QY 169 VISTNDQILHNVAQIAH-CYAKKIGSGFDVATAIYGLIYRRFPALINDVQVLESPP 227
 Db 382 VLNKLDLRLRYKAYVYKKKAQE-----KGLP-----EPTVFN----- 418
 QY 228 EKFEPTLKKLIESNWEKHERCTLPGIKILMGDVKGSETPPLVSRVLQWKKE----- 281
 Db 419 EEEVEELKKGIL-----LDMGIKLTFKVK--SLTKLKKKIPKKEDNQAVD 464
 QY 282 -----KPESSVYVDQNSANLQ---FMKLR----- 306
 Db 465 TKSMEEPKVAQOPALRGVEPTEDSNINMSINNVMDEIDPEKELINNNTPNVVPTQSK 524
 QY 307 -----MRKXVSDPEYIKE-----LDHSVEPLVAIK-----NIRKG 339
 Db 525 KKKKNETVSGMDENFDNHPENYKKEYYIDNDMEVKAKKIGVTLKKEPLKNGVSET 584
 QY 340 LQAL-----TQKSEVPIEPDVQTLDRCOEIPGCVGVPGAGYDAIATVLENQVGN 394
 Db 585 IKLIHGNKDKKHIEAIINNDIQIK---QELQAIYNELMNTNNGKNIQOI----- 632
 QY 395 FKQKTEPNDFYFNNAVYVWLEETEGVLE 424
 Db 633 PQONILN-DVNLQETEEEMEKQVEATIQ 661

RESULT 38

JX0340
 gramicidin S synthase 2 - Bacillus brevis
 N/Alternate names: gene Grs2 protein
 C/Species: Bacillus brevis
 C/Date: 22-Apr-1995 #sequence_revision 26-May-1995 #ext_change 03-Nov-2000
 C/Accession: JX0340
 R/Saito, F.; Hori, K.; Kanda, M.; Kurotsu, T.; Saito, Y.
 J. Biochem. 116, 357-367, 1994
 A/Title: Entire nucleotide sequence for Bacillus brevis Nagano Grs2 gene encoding gramicidin S synthase
 A/Reference number: JX0340; MUID:95122465; PMID:7822255
 A/Accession: JX0340
 A/Molecule type: DNA
 A/Residues: 1-4450 <SAT>
 A/Cross-references: DDBJ:D29676; NID:G473695; PIDN:BA06146.1; PID:G511490
 C/Comment: This enzyme catalyzes activation and combination of four constituent amino ad

C/Comment: This enzyme contains 4 domains with a mean of 1042 amino acid residues contain
 synthase 1.
 C/Suprafamily: alpha-aminoacyl-L-cysteine-L-valine synthetase; acetate-CoA ligase homolog
 C/Keywords: carrier protein; phosphopantetheine; phosphoprotein
 F/514-956/Domain: acetate-CoA ligase homolog <ACLI>
 F/514-1042/Domain: acyl carrier protein homolog <ACP1>
 F/1552-1991/Domain: acetate-CoA ligase homolog <ACI2>
 F/2009-2077/Domain: acyl carrier protein homolog <ACP2>
 F/2588-3036/Domain: acetate-CoA ligase homolog <ACI3>
 F/3054-3122/Domain: acyl carrier protein homolog <ACP3>
 F/3634-4074/Domain: acetate-CoA ligase homolog <ACI4>
 F/4092-4160/Domain: acyl carrier protein homolog <ACP4>
 F/1006,2041,3086,4124/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 4.8%; Score 107.5; DB 2; Length 4450;
 Best Local Similarity 19.4%; Pred. No. 2,5e+02;
 Matches 69; Conservative 51; Mismatches 140; Indels 95; Gaps 12;

QY 36 HAVTPKGTSLKESRIKISSPOFANGEWEMHISN-----TEKREV----- 77
 Db 3604 HIAIVWEGALVYHNLKANDLAVLEKGVTPHVAIMTERSELMIVGIFSLKAG 3663
 QY 78 -----QSRNPFLE---ATFIVLAIQPTKADLEIIITSDPECHSQEDTEKT 124
 Db 3664 AVVPIDPAVYPERIQLYLEDSCATLLQSHVNLKLPVDIEWLDLTDEQNVYEDGTNLP 3723
 QY 125 SNGEFTLYHSRAITEVEKTLGSSAGLVSVATSLSHFIENYISTNDQILHNVAQIA 184
 Db 3724 -----MNOSTDLAIITYS-----GTYGPKGMIMHOSII 3754
 QY 185 HCYAKK---IGSGFDVATAYGL---IYRRFPALINDVQVLESDEPKFTELEK 236
 Db 3755 NCLQNRKEBERGPG-DTALQVSPAFDGFVSLPAPILAGATSVLPKEEKADPVALLK 3813
 QY 23 LIESNWEKHERCTLPYGIKILMG---DYK-----GSEFPKLVSRVLQWK 280
 Db 3814 LIAS-----BEITHYVGPSPFSAIILVSSKDLQNLRCYTLGSEKLPQAVIKKIKERN 3867
 QY 281 EKPESSVYVDQNSANLQFMKELREMEKEDSDP---ETVYKELDSHVEPLV 331
 Db 3868 KEIENVNEXGPTENSVVTIMRDIVQEQEITIGCLSNVDYIIVNCHQLOPQGV 3922

RESULT 39

S67593
 transport protein USO1 - yeast (Saccharomyces cerevisiae)
 N/Alternate names: protein D2552; protein YDL058w
 C/Species: Saccharomyces cerevisiae
 C/Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #ext_change 21-Jul-2000
 C/Accession: S67593; A38455; S30782
 R/Bloeker, H.; Brandt, P.
 submitted to the Protein Sequence Database, July 1996
 A/Reference number: S67587
 A/Accession: S67593
 A/Molecule type: DNA
 A/Residues: 1-1790 <BLQ>
 A/Cross-references: EMBL:Z74106; NID:G1431058; PID:G253003; PID:G1431059; MIPS:YDL058w
 A/Experimental source: strain S288C
 R/Nakajima, H.; Hirata, A.; Ogawa, Y.; Yonehara, T.; Yoda, K.; Yamaoka, M.
 J. Cell Biol. 113, 245-260, 1991
 A/Title: A cytoskeleton-related gene, USO1, is required for intracellular protein transp
 A/Reference number: A38455; MUID:91185402; PMID:2010462
 A/Accession: A38455
 A/Molecule type: DNA
 A/Residues: 1-389, 'TA', 392-724, 'S', 726-1790 <NAK>
 A/Cross-references: GB:X54378; NID:G47777; PIDN:CA43253.1; PID:G4778
 A/Note: The authors translated the codon ACT for residue 768 as Ile
 R/Hosletter, M.R.; Hexman, D.J.; Bender, C.M.; McClellan, M.; Tao, N.; Kendrick, K.E.
 submitted to the EMBL Data Library, February 1993
 A/Description: An integrin analogue in Saccharomyces cerevisiae.
 A/Reference number: S30782
 A/Accession: S30782
 A/Molecule type: DNA

A:Residues: 71-846, 'E', 848-923, 'K', 925-1252, 'I', 1254-1318, 'V', 1320-1460, 'S', 1462-1580, 'S'
 A:Cross-References: EMBL:L03188
 C:Genetics:
 A:Gene: SGD:USO1; INT1
 A:Cross-References: SGD:S0002216; MIPS:YDL058w
 A:Map position: 4L
 C:Keywords: coiled coil; transmembrane protein
 F:326-342/Domain: transmembrane #status predicted <TM1>
 F:394-410/Domain: transmembrane #status predicted <TM2>
 F:617-633/Domain: transmembrane #status predicted <TM3>

Query Match 4.8%; Score 107; DB 2; Length 1790;
 Best Local Similarity 20.2%; Pred. No. 68;
 Matches 87; Conservative 67; Mismatches 166; Indels 110; Gaps 16;

```

QY 13 LAGYLVEPIVDAYVTALSSRMHAYITPKTSLKSRKISSPOFANGEMWYHISNTE 72
DB 967 LANNYKDMOENESLIKAVEE-----SKNESSIQLSNLQ-----NKIDSMSQ 1008
QY 73 KPREVQSRINPFLKATIFVLAVIOTPEAFDEIITISDPG---YHSQ-----EDTETKT 124
DB 1009 EKENPQIE-RGSIENKIEQLKKTISDLQETKEIITISKSDSKDYESQISLKEKLETTAT 1067
QY 125 SNGEKTFLYHRAITEVEKTLGSSAGLVSVVATSLSHFLPNVISTNKDILHNVAQIA 184
DB 1068 TANDENV-----NKISELTKEEELAEI---AAYNLKNELETKEETSEKALKEVENE 1119
QY 185 HCYAKKTISSGFDAVTAIYGLVYRFPALINDVQVLESDEPEFTELKLIIE--SNW 242
DB 1120 EHLKEEKTQLEKEATET-----KQILNSIPANLESLEKEHEDLAQLKYEEOIANK 1171
QY 243 EKHHERCTLPYGIKILMDVKGSGS-----TPKLVSRVLQWKKKEPSSVYVDQAN 294
DB 1172 EQQYNE-----EISQINBETISTQGENESIKKKNELBESEVAMKSTSEQSNLKKSEHD 1226
QY 295 SANLQPMKELRMREKTYSDPEYIKELDHSVEPLTVAIKNIRKGLQALTOKSEVPIEPD 354
DB 1227 ALNLQ-----IKELKKNETN-EASLIESIKSYSESETVKIK-----1261
QY 355 VQTQLDRCQETPGCVGVVPGAGYDAIAVLVLENQVGNFKQKTLNPDYFHHVYVDL 414
DB 1262 ---ELQDECN-----FKEKEVSELDKAKASD--KNSKYLEL 1294
QY 415 EEQTEGVLEE 424
DB 1295 QKSEKIKEE 1304

```

RESULT 40

G70163
 hypochetrical protein BB0512 - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
 C:Accession: G70163
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Jathigra, R.; White
 son, D.; Peterson, J.; Karlavag, J.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
 ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horek, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943; PMID:9403685
 A:Accession: G70163
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2166 <KLB>
 A:Cross-References: GB:AE001153; GB:AE000783; NID:92688419; PIDN:AA66876.1; PID:9268842
 A:Experimental source: strain B31

Query Match 4.8%; Score 107; DB 2; Length 2166;
 Best Local Similarity 18.6%; Pred. No. 91;
 Matches 85; Conservative 62; Mismatches 169; Indels 142; Gaps 16;

QY 44 TSLKSRIKISSPOFANGEMWYHISNTEKPREVQSRINPFLKATIFVLAVIOTPEAFD 103

```

DB 616 TOLRESKIQYS-----KMOGEMDSNL---KNIESQINKTWEE--FLSLIQIQDKGIE 663
QY 104 LEIIVSPGVHSQEDTEFTKTSNGEKTFVHRAITEVEKTLGSSAGLVSVVATSLIS 163
DB 664 LSESVFNLSDBHIQ-----KKAIDMGGS--WDELIANKSLDIDVSSBELS 710
QY 164 HPIPNVISTNKDILHN-----AQIAHCYAKKTISSGFDAVTAIYGLVYRFPQ 212
DB 711 SATLKIESLEKDVNDRMREYVLLKTDISLVIEKTELKMDSYSGSDAELIGKPFIRNQ 770
QY 213 PALIND--VFOV-----LESQP--EKPFTELKLIIE 239
DB 771 TEIIRKQSVFMLEDLNKKFDDKNNFVISKIECDYKLDQFKIESDILNFFKSDINPFI 830
QY 240 -----SWEEKHERCTLPYGIK 256
DB 831 SKLQIVSNIKSDNOKQIDFLDRISKDILNKRDSINNEVDSKLSDMQSKLMEITVAKIENL 890
QY 257 LLMG-----DVKGSETPKL-----VSRVLQWKKKEPPE--SSVY--DQILNSANLQFM 301
DB 891 LSSGVNDDLDISEVTTIKELKTSIESLESYIEKIDEPNQAAYISDELQDILNHFN 950
QY 302 KELRMREKYP-----SDPEYIKELDHSVEPLTVAIKNIRKGLQALTOKSEVPIEPD 354
DB 951 KETRELLENLSKFAAVLANSEEFVKEVDSLLQDKRTDIASFQANIDITLDSLVNKFN-- 1008
QY 355 VQTQLDRCQETPGCVGVVPGAGYDAIAVLVLENQVGNFKQKTLNPDYFHHVYVDL 414
DB 1009 -----DINKINGKYEVIISNYRGYSNISKLENEI 1040

```

Search completed: August 8, 2003, 09:40:32
 Job time : 49 secs

XX WPI; 2001-218441/22.
DR N-PSDB; AAD02791, AAD02792.
XX
XX
PT New polypeptides and polynucleotides (ERGs) from *Candida albicans*,
PT useful in assays for identifying inhibitors of phosphomethyltransferase
PT activity and as reagents for diagnosing *C. albicans* infection -
XX
PS Claim 1; Page 27-28; 29pp; English.
XX
XX The present sequence is phosphomethyltransferase kinase (PMK; ERG) protein
CC from *Candida albicans*. The ERG protein is useful in an assay for
CC identifying compounds that inhibit phosphomethyltransferase kinase (PMK)
CC activity. These inhibitors are useful as anti-fungal agents. The ERG
CC DNA and protein are also useful as reagents for diagnosing *C. albicans*
CC infection.
XX
SQ Sequence 432 AA;
XX
Query Match 100.0%; Score 2230; DB 22; Length 432;
Best Local Similarity 100.0%; Pred. No. 2, 9e-196;
Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MSKAFSAPGKAFIAGGYLVLEPIYDAVYALSSRMHNAVTPKTSIKESRIKISSPOFAN 60
1 MSKAFSAPGKAFIAGGYLVLEPIYDAVYALSSRMHNAVTPKTSIKESRIKISSPOFAN 60
DB
QY 61 GEWEYHISSTNTEKPREVQSRINPFLAATIFIVLAYIQTEAPDLIIIIYSDPGYHSQEDT 120
61 GEWEYHISSTNTEKPREVQSRINPFLAATIFIVLAYIQTEAPDLIIIIYSDPGYHSQEDT 120
DB
QY 121 ERTSSNGEKTFLYHSRAITEVEKTLGSSAGLVSVATSLSHFIPNVISTNKDILHNV 180
121 ERTSSNGEKTFLYHSRAITEVEKTLGSSAGLVSVATSLSHFIPNVISTNKDILHNV 180
DB
QY 121 ERTSSNGEKTFLYHSRAITEVEKTLGSSAGLVSVATSLSHFIPNVISTNKDILHNV 180
121 ERTSSNGEKTFLYHSRAITEVEKTLGSSAGLVSVATSLSHFIPNVISTNKDILHNV 180
DB
QY 181 AQIAHCYAQKKIGSGDFVATAIYGLIVRRFOPALINDVFOVLESDEKPETELKILIES 240
181 AQIAHCYAQKKIGSGDFVATAIYGLIVRRFOPALINDVFOVLESDEKPETELKILIES 240
DB
QY 241 NMEKHERCTLPYGIKILMDVKGSGSTPLVSRVLOMKKEKPEBSVVYDQNSANLQF 300
241 NMEKHERCTLPYGIKILMDVKGSGSTPLVSRVLOMKKEKPEBSVVYDQNSANLQF 300
DB
QY 301 MKELREMERKYSDDPEYIKELDSVEPLTVAIKNIRKGLQALTOKSEVPIEDVOTQL 360
301 MKELREMERKYSDDPEYIKELDSVEPLTVAIKNIRKGLQALTOKSEVPIEDVOTQL 360
DB
QY 361 DRCOEIPGCVGVPPAGGYDAIAVLVLENOGVNFQKLTLENDYFHNVYVWDLSEQTEG 420
361 DRCOEIPGCVGVPPAGGYDAIAVLVLENOGVNFQKLTLENDYFHNVYVWDLSEQTEG 420
DB
QY 421 VLEKPEDYIGL 432
421 VLEKPEDYIGL 432
DB

RESULT 2

AAU15094 standard; Protein; 432 AA.
ID AAU15094 standard; Protein; 432 AA.

XX
XX
AC AAU15094;
XX
DT 04-DEC-2001 (first entry)
XX
DE Protein encoded by *C. albicans* essential gene *CayMR220W* (ERG8).
XX
XX Gene identification; essential gene; GRACE; pathogenic fungus;
KM gene replacement and conditional expression; fungal infection.
XX
OS *Candida albicans*.
XX
XX WO200160975-A2.
XX

PD 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US05551.
XX
XX 18-FEB-2000; 2000US-0183534.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H;
XX
XX WPI; 2001-489080/53.
DR N-PSDB; AAS23422.
XX
PT Identifying genes essential to fungal metabolisms and identifying
PT potential therapeutic agents that target these genes -
XX
PS Claim 43; Page 226-227; 324pp; English.
XX
XX The present invention relates to novel methods for constructing fungal
CC strains useful for identification and validation of gene products as
CC targets for therapeutic agents, for creating a collection of identified
CC essential genes, and screening assays for the discovery of new drugs.
CC The invention provides the GRACE (gene replacement and conditional
CC expression) method for the construction of mutant organisms referred to
CC as GRACE strains of the organism. The invention can be applied to any
CC organism, particularly a pathogenic fungus e.g. *Candida albicans*,
CC *Aspergillus fumigatus* and *Cryptococcus neoformans*. The methods are
CC useful to identify agents that may be used in the treatment of fungal
CC infections. AAU15053-AAU15113 represent proteins encoded by *C. albicans*
CC essential genes.
XX
SQ Sequence 432 AA;
XX

Query Match 98.7%; Score 2202; DB 22; Length 432;
Best Local Similarity 98.8%; Pred. No. 1, 1e-193;
Matches 427; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKAFSAPGKAFIAGGYLVLEPIYDAVYALSSRMHNAVTPKTSIKESRIKISSPOFAN 60
1 MSKAFSAPGKAFIAGGYLVLEPIYDAVYALSSRMHNAVTPKTSIKESRIKISSPOFAN 60
DB
QY 61 GEWEYHISSTNTEKPREVQSRINPFLAATIFIVLAYIQTEAPDLIIIIYSDPGYHSQEDT 120
61 GEWEYHISSTNTEKPREVQSRINPFLAATIFIVLAYIQTEAPDLIIIIYSDPGYHSQEDT 120
DB
QY 121 ERTSSNGEKTFLYHSRAITEVEKTLGSSAGLVSVATSLSHFIPNVISTNKDILHNV 180
121 ERTSSNGEKTFLYHSRAITEVEKTLGSSAGLVSVATSLSHFIPNVISTNKDILHNV 180
DB
QY 121 ERTSSNGEKTFLYHSRAITEVEKTLGSSAGLVSVATSLSHFIPNVISTNKDILHNV 180
121 ERTSSNGEKTFLYHSRAITEVEKTLGSSAGLVSVATSLSHFIPNVISTNKDILHNV 180
DB
QY 181 AQIAHCYAQKKIGSGDFVATAIYGLIVRRFOPALINDVFOVLESDEKPETELKILIES 240
181 AQIAHCYAQKKIGSGDFVATAIYGLIVRRFOPALINDVFOVLESDEKPETELKILIES 240
DB
QY 241 NMEKHERCTLPYGIKILMDVKGSGSTPLVSRVLOMKKEKPEBSVVYDQNSANLQF 300
241 NMEKHERCTLPYGIKILMDVKGSGSTPLVSRVLOMKKEKPEBSVVYDQNSANLQF 300
DB
QY 301 MKELREMERKYSDDPEYIKELDSVEPLTVAIKNIRKGLQALTOKSEVPIEDVOTQL 360
301 MKELREMERKYSDDPEYIKELDSVEPLTVAIKNIRKGLQALTOKSEVPIEDVOTQL 360
DB
QY 361 DRCOEIPGCVGVPPAGGYDAIAVLVLENOGVNFQKLTLENDYFHNVYVWDLSEQTEG 420
361 DRCOEIPGCVGVPPAGGYDAIAVLVLENOGVNFQKLTLENDYFHNVYVWDLSEQTEG 420
DB
QY 421 VLEKPEDYIGL 432
421 VLEKPEDYIGL 432
DB

RESULT 3

ABP73186 standard; Protein; 432 AA.
ID ABP73186 standard; Protein; 432 AA.

ABP73186;
30-JAN-2003 (first entry)
Candida albicans essential protein SEQ ID NO 7023.
Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis; signal transduction; DNA replication; cell division; growth; proliferation; Candida albicans; fungicide; antifungal.
Candida albicans.
WO200253728-A2.
11-JUL-2002.
26-DEC-2001; 2001WO-US49486.
29-DEC-2000; 2000US-259128P.
20-FEB-2001; 2001US-079202A.
22-AUG-2001; 2001US-314050P.
(ELIT-) ELITRA PHARM INC.
Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
WPI; 2002-566594/60.
N-PDSB; AB231736.
Constructing strains for identifying gene products as effective targets for therapeutic intervention, by inactivating in the strain one allele of a gene and placing other allele of the gene under conditional expression
Claim 44; SEQ ID NO 7023; 167pp + Sequence Listing; English.
The invention relates to constructing (M1) a strain of diploid fungal cells in which both alleles of a gene are modified, comprising modifying one allele by insertion or replacement by a cassette having an expressible selectable marker and modifying other allele by recombination, of a promoter replacement fragment with a heterologous promoter, so that expression of the second allele is regulated by the promoter. (M1) is useful for constructing a strain of diploid fungal cells in which both alleles of a gene are modified. The diploid fungal cells having both alleles modified are useful for identifying a gene that is essential to the survival or growth of a fungus, a gene that contributes to the virulence and/or pathogenicity of a fungus, a gene that contributes to the resistance of a diploid fungus to an antifungal agent, an antifungal agent that inhibits the growth of a diploid fungus and for identifying a therapeutic agent for treatment of a mammalian disease. (M1) is useful for identifying a compound which modulates the activity of a gene product, preferably enzymatic activity, carbon compound catabolism, biosynthetic, transporter, transcriptional, translational, signal transduction, DNA replication and cell division activity. The method is useful for identifying a compound having the ability to inhibit growth or proliferation of C. albicans cells and for treating infection by C. albicans. The present sequence is that of an essential Candida albicans protein used in the method of the invention. Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to Derwent by the European Patent Office.

Query Match	98.7%	Score 2202	DB 23	Length 432
Best Local Similarity	98.8%	Pred. No. 1.1e-193		
Matches 427; Conservative	2;	Mismatches 13;	Indels 0;	Gaps 0;

[illegible]

QY	61	GEMEHHSNNTEKPEVOSRINPELEATFIYLAIOPTAEADLXIIISDPECHSOEPT	120
Dp	61	GEMEHHSNNTEKPEVOSRINPELEATFIYLAIOPTAEADLXIIISDPECHSOEPT	120
QY	121	ETKTSNNEKTFLYHSRAITTEVEKTKGLSSAGLVSVVATSLSHFIPNVISTKDKILHNV	180
Dp	121	ETKTSNNEKTFLYHSRAITTEVEKTKGLSSAGLVSVVATSLSHFIPNVISTKDKILHNV	180
QY	181	AOIAHCAOQKIGSGFDVATAIYGLIVYRRFOPALINDVFOVLESDDPEKFPTELKULIS	240
Dp	181	AOIAHCAOQKIGSGFDVATAIYGLIVYRRFOPALINDVFOVLESDDPEKFPTELKULIS	240
QY	241	NMEEKHERCTLPYGIKILMGDVKGSGSETPKLYSRVLQWKKKPESSVYVYDOLNSANLAF	300
Dp	241	NMEEKHERCTLPYGIKILMGDVKGSGSETPKLYSRVLQWKKKPESSVYVYDOLNSANLAF	300
QY	301	MKELEHREMEKUDSPETIYIKELIDHSVEPLTVAIKNIRKGLQALTOKSEVPIEBDVOTOLL	360
Dp	301	MKELEHREMEKUDSPETIYIKELIDHSVEPLTVAIKNIRKGLQALTOKSEVPIEBDVOTOLL	360
QY	361	DRCOEIPBCGVGVPGAGGYDAIAVLVLBNQVGNFKOTLENPDYFHHNYWMDLEBQOTGS	420
Dp	361	DRCOEIPBCGVGVPGAGGYDAIAVLVLBNQVGNFKOTLENPDYFHHNYWMDLEBQOTGS	420
QY	421	VLEEKPEPDYIGL 432	
Dp	421	VLEEKPEPDYIGL 432	

RESULT 4
ABP53757
ID ABP53757 standard; Protein; 451 AA

DT 03-JAN-2003 (first entry)

DE *Saccharomyces cerevisiae* mevalonate kinase SEQ ID NO:24.
XX
KW Mevalonate kinase; enzyme; herbicide; agricultural; plant

PN	US2002119546-A1.
XX	
PD	29-AUG-2002.

PF 20-JUL-2001; 2001US-0909745.

PR	05-NOV-1998;	98US-107241P.
PR	04-NOV-1999;	99US-0433242.

PA (FALC/) FALCO S. C.
PA (FAMO/) FAMODU O. O.

PI	Falco SC, Famodu OC
XX	
DR	WPI; 2002-731357/79.

PT New mevalonate kinase polypeptide useful for new herbicide discovery
PT and design, and for altering levels of mevalonate kinase in transformed
PT cells -

PS Example 4; Fig 3; 39pp; English.

CC The present invention describes a isolated mevalonate kinase polypeptide
CC (I), having a sequence identity of 80 % based on the Clustal method of
CC alignment when compared to corn, rice, soybean and wheat mevalonate
CC kinase. (I) has herbicide and agricultural activities. A nucleic acid
CC (II) encoding (I) is useful for transforming a cell and producing a
CC transgenic plant. A chimeric gene (III) comprising (II) is useful for
CC altering the level of expression of a squalene biosynthetic enzyme in a
CC host cell, where the expression of the chimeric gene results in

DB 291 LEMRKQNOEADMLWALQSNNEERLLEARRLAQSPD-----EHTLSDFFENVRYY 340
QY 333 IKNIKRGALATQKSEVPIEPDVQOTLDRCOEIPCCGCVWPGAGYDAIVLENO- 391
DB 341 IORSRHHISMVQKSDVPIEPVQTELLALSLBSGVIGVPGAGYDAIVLLIQDNPD 400
QY 392 -VGNFKQ--KTLEN---PDY---FHNVWVWDLSEQTEGVLEKPEDEYIG 431
DB 401 VTRLKAFPETWESKAEDDFGKIGKVRLLGVRRHSGEVKNEMLEQYAG 449
RESULT 6
ABJ26348 standard; Protein; 484 AA.
AC ABJ26348;
XX 16-APR-2003 (first entry)
XX Aspergillus fumigatus essential gene protein #1006.
DE Fungicide; cyrostatic; essential gene; Aspergillus fumigatus; infection;
XX cancer; contamination; biofilm; antibody; immune response.
OS Aspergillus fumigatus.
XX WO200266090-A2.
XX 31-OCT-2002.
XX 23-APR-2002; 2002WO-US13142.
XX 23-APR-2001; 2001US-285697P.
XX 27-APR-2001; 2001US-287066P.
XX 05-JUN-2001; 2001US-295890P.
XX 09-JUL-2001; 2001US-303899P.
XX 31-AUG-2001; 2001US-316362P.
XX (ELIT-) ELITRA PHARM INC.
XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM,
PI WPI; 2003-093124/08.
DR New purified or isolated nucleic acids of essential genes of
PT Aspergillus fumigatus, useful for treating or preventing infections by
PT A. fumigatus, or for treating a non-infectious disease in a subject
PS e.g. cancer -
XX Disclosure; Page -; 175pp; English.
XX The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively

CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case of virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention.
XX
SQ Sequence 484 AA;
Query Match 29.4%; Score 655.5; DB 24; Length 484;
Best Local Similarity 33.1%; Pred. No. 2.9e-51;
Matches 166; Conservative 81; Mismatches 152; Indels 103; Gaps 16;
QY 4 AFSAPGKAPLAGVLEPIYDAVYVTLSSRMHVI-----TFK 42
DB 10 ALSAPGKVLTCGYLVDRSYTGTFALDARHIVVQGLRRHRRRAASGASHGSDPQ 69
QY 43 -----GTSLKESRIKISSPOFANGSEYHI-----SNTEKPEVQSRINPL 85
DB 70 AEGNVHGDKDEGTIVHSPQFVDVWVEYSIORCEDGGVLYKQRNDPR-----NLFV 123
QY 86 EATIRIVLAYI---QPEAPDLEIIISDGYHSQEDTERTTSNGKRTPLHSHRAITE 141
DB 124 ETSINPALTYSYVADSKDFGSLSTITLADNDYSETAFSKASGLRSSRFVDFGVRLQE 183
QY 142 VEKTLGSSAGIVSVATSLSH--FIPNVISTNKDILHNVQAIAHYAOKKIGSGFDVA 199
DB 184 AHKTGLSSAALVTLVSSLVHRTMQPDDLGPGRDKLHNLAAHCAAGKVGSGFDVA 243
QY 200 TAIYGLIYRRPQALINDV-----FOVLESDEPEKFPTELKLISSNNEKHE 247
DB 244 AAIYSGCLYRRFSPILHSVGDAGSPGEERLFRIVEDADPQH-----WDTE-- 291
QY 248 RC-----TLPGIKILMDGVGSGSETPKLYSRVLOMKKEKESVVYDQINSANLQPMK 302
DB 292 -CLDFGKMLPRGMQVNLCDVBCGSGTPEMVKRVLEADMLWALQSNNEERLRL 350
QY 303 ELREMERKDYSDPEYIKELDHSV--EPLTVAIKNIKGLATQKSEVPPIEPVQOTL 359
DB 351 ELRLAQSPD-----EHTLSDFFENVRYYIORSNHHISMVQKSDVPIEPVQTEL 400
QY 360 LDRCEIIPGCVGVVPGAGYDAIVLENO--VGNFKQ--KTLEN---PDY---FHNV 409
DB 401 LDALSELBGVIGVVPAGYDAIVLLIQDNPDVTRLKAFPETWESKAEDDFGKIGV 460
QY 410 YWVDEBQTEGVLEKPEDEYIG 431
DB 461 RLGVRRHSGEVKNEMLEQYAG 482
RESULT 7
ABB77505
ID ABB77505 standard; Protein; 505 AA.
XX ABB77505;
XX 29-JUL-2002 (first entry)
XX Arabidopsis thaliana pmwK SEQ ID NO 2.
XX Thale cress; pmwK; phosphomevalonate kinase; plant; herbicide;
XX growth regulator; enzyme.
XX Arabidopsis thaliana.
XX DE10057755-A1.
XX 23-MAY-2002.
XX 22-NOV-2000; 2000DE-1057755.
XX 22-NOV-2000; 2000DE-1057755.
XX (FARB) BAYER AG.

XX Meisner R, Lechelt-Kunze C;
 XX WPI, 2002-445360/48.
 DR N-PSDB; ABL60244.
 XX
 PT New nucleic acid encoding plant phosphomevalonate kinase, useful for
 PT identifying modulators, potentially useful as herbicides and growth
 PT regulators
 XX
 PS Claim 14, Page 13-14, 18pp; German.
 XX
 CC The invention relates to a nucleic acid (1, ABL60244) that encodes a
 CC plant phosphomevalonate kinase (PMVK, AB877505), excluding the known
 CC fully defined partial sequences ABL60245-ABL60247. Plant PMVK (1),
 CC constructs and host cells that contain (1) are used to identify agents
 CC that bind to and/or modulate activity of PMVK, potentially useful as
 CC herbicides and growth regulators. (1) is also used for recombinant
 CC production of PMVK.
 CC
 XX
 SQ Sequence 505 AA;
 Query Match 21.3%; Score 475; DB 23; Length 505;
 Best Local Similarity 28.3%; Pred. No. 1.3e-34;
 Matches 143; Conservative 78; Mismatches 184; Indels 100; Gaps 15;
 QY 1 MSKAFAPGKAFIAGYLVLEPIYDAYVTALSSRMNAVITPKGTSLKE-----SRIKI 53
 DB 1 MAVVASAPGKVLMTGGVILEKPNAGVLSTNARFAYIVPINEEVPESWAMKMTDVKL 60
 QY 54 SSPQFANGWEYHISNT--TEKPREVOSRINPELEATIFVLAYIOPTFAFD----- 103
 DB 61 TSPQJLR-ESMYKLSLNHLTLQSVASDSNRPVEHAIQVAAIAAHILATEKDESLAKLL 119
 QY 104 ---LEIITVSDPGYHSQED-----TETKTSNGEKTFYHRS 137
 DB 120 LQGLDITILGSDNPFYSRNQIESAGLPLTPESIGTLAPFASITFNAAESNG-----AN 172
 QY 138 AITEVEKTLGSSAGLVAVATSLSHFIPNVIS-----TNDIILHNVAAQIAH 185
 DB 173 SKREVAKTGLSSAAMTTAVAAALL-HYL-GVVDLSDPCKEKGFGSDLDVHIMIQTSH 230
 QY 186 CVAKKIGSGFDVATAIYGLIVRRFOPALINDVFOVLESBP--EKFPTELKILIESNME 243
 DB 231 CLAQKGVSGFVDSCAVYSGQRVRSPEVLSFAQVAVIGLPLNEVIGITLK-----GKMD 286
 QY 244 EKHERCTLPYGIKLMGDV-KGSETPKLVSRVLQWKEKEPSSVAVYDQNSANTQPMK 302
 DB 287 NKRTBFLPLMNLFLGEPGSGSSTPSWVGAVKKQMSPERKARENQVLS DANLELET 346
 QY 303 ELREMEKXDSDEYIKELD-----HSEVPLTVAL-----KNIRKG 339
 DB 347 KANDLSKLAKDHDVLRVTKSCSVITSEKVLHAPEINELAIKELLEAREMLRIRLL 406
 QY 340 LQATLTKSEVLEPDPVOTGLDRCOIRPGCVGVGVAGGYDAIAVLVLENQVGNFKQKT 399
 DB 407 MRQMGMAAIVPIREPESQGLDSTMSAEGVLLAGVGAGGFDAIFAITLGDSTKLTQAM 466
 QY 400 LENPDYFHNVYVWDLBEQTEGVLEE 424
 DB 467 SS-----HNVLTALLVREDPHGVCLF 486
 RESULT 8
 ID AAO27025 standard; Protein; 503 AA.
 XX
 AC AAO27025;
 XX
 DT 15-MAY-2003 (first entry)
 XX
 DE Acetyl-coA acetyltransferase enzyme protein, SEQ ID No 12.
 XX

KM Isopentenyl diphosphate; IPP; pathway enzyme; IPP biosynthesis;
 KM acetyl-coA acetyltransferase enzyme; acetate.
 XX
 OS Hevea brasiliensis.
 XX
 PN WO2003010294-A2.
 XX
 PD 06-FEB-2003.
 XX
 PF 23-UTL-2002; 2002WO-US24048.
 XX
 PR 25-UTL-2001; 2001US-307673P.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Hallahan DL, Keiper-hrynko NM;
 DR WPI; 2003-239439/23.
 DR N-PSDB; AAL55372.
 XX
 PT Novel isolated nucleic acid molecule encoding isopentenyl diphosphate,
 PT IPP, pathway enzyme, useful for obtaining nucleic acid molecule
 PT encoding IPP pathway enzyme, and for regulating IPP biosynthesis in
 PT organism
 XX
 PS Claim 4; Page 62-63; 66pp; English.
 XX
 CC This polynucleotide sequence represents an isolated nucleic acid molecule
 CC which encodes an isopentenyl diphosphate (IPP) pathway enzyme that has a
 CC 411, 464, 386, 503 or 415 residue amino acid sequence, given in the
 CC specification, hybridizes with nucleic acid molecule encoding the amino
 CC acid sequences, or is complementary to the sequences. The isolated
 CC nucleic acid is useful for regulating IPP biosynthesis in an organism,
 CC where the nucleic acid is overexpressed such that IPP biosynthesis is
 CC altered in the organism. The IPP pathway gene is over-expressed on a
 CC multicopy plasmid, and is operably linked to an inducible or regulated
 CC promoter. The IPP gene is optionally expressed in antisense orientation,
 CC or is disrupted by insertion of foreign DNA into the coding region. The
 CC isolated IPP nucleic acid or sequences showing identity are useful for
 CC obtaining nucleic acid molecules encoding IPP pathway enzymes, which
 CC involves probing a genomic library with the nucleic acid, identifying a
 CC DNA clone that hybridizes with the nucleic acid, and sequencing the
 CC genomic fragment that comprises the clone, where the sequenced genomic
 CC fragment encodes an IPP pathway enzyme. The isolated nucleic acid having
 CC a 1233, 1392, 1158, 1509 or 1245 nucleotide sequence, given in the
 CC specification, is useful for obtaining a nucleic acid molecule encoding
 CC an IPP pathway enzyme, which involves synthesizing at least one
 CC oligonucleotide primer corresponding to a portion of the sequence, and
 CC amplifying an insert present in a cloning vector using the
 CC oligonucleotide primer, where the amplified insert encodes a portion of
 CC an amino acid sequence encoding the enzyme. A transformed host cell is
 CC useful for producing a compound in the IPP pathway, which involves
 CC contacting a transformed host cell transformed with the isolated IPP
 CC nucleic acid under the control of suitable regulatory sequences, under
 CC suitable growth conditions with a carbon substrate, thus a compound in
 CC IPP pathway is produced. This sequence represents an acetyl-coA
 CC acetyltransferase enzyme protein, a variant of one of the enzymes used to
 CC synthesize IPP from acetate.
 XX
 SQ Sequence 503 AA;
 Query Match 20.0%; Score 445.5; DB 24; Length 503;
 Best Local Similarity 27.4%; Pred. No. 6.5e-32;
 Matches 138; Conservative 88; Mismatches 179; Indels 99; Gaps 16;
 QY 1 MSKAFAPGKAFIAGYLVLEPIYDAYVTALSSRMNAVITPKGTSLKE-----SRIKI 53
 DB 1 MAVVASAPGKVLMTGGVILEKPNAGVLSTNARFAYIVPINEEVPESWAMKMTDVKL 60
 QY 54 SSPQFANGWEYHISNT--TEKPREVOSRINPELEATIFVLAYIOPTFAFD----- 103
 DB 61 TSPQJLR-ESLYKLSLNHLTLQSVASDSNRPVEHAIQVAAIAAHILATEKDESLAKLL 119

```
QY 104 ----LEIITYSDPGHSD-----TETKTSNCKEPTLYHS 136
DB 120 LLOGLDITLTGNDPYSYNEIEACGLPTPEISALPFSSITFVNEBANGQ----- 172
QY 137 RAITEVEKGLSSALVSVATSLSHF-IRPVISTNK-----DILHNAQIACHYA 188
DB 173 NCKPEVAKTGLSSAMMTAAVAALLHGLVLDSSCKEKRSDLDVHIIAQTHACTA 232
QY 189 OKKIGSGFVATAIYGLIYRRFQPALINDVFOVLESDEKPFTELKLIESNWEKHER 248
DB 233 QOKVSGGFVSSANVSHRYRSPREVLSADAGGIP--LQEVSNILKGMKHERIM 290
QY 249 CTLPYGIKILMGDV-KGSEETPLVSRVLQMKKEKESSVYVQIUNAN-----LQPM 301
DB 291 FSLPLMSILIGBPGSGSTPSMGALKKWQKSDTOKSOETWRKLSANSLALETQFNIL 350
QY 302 KKLREKREKYS-----DEPTYIKELDHSVEPLTVAI-----KNIRKGLQA 342
DB 351 SKLAE--EHMDAYKVIDSCSTKNSKWEIQATEPEREAVVALLGSRNAMLIQIRNYMQ 408
QY 343 LTQKSEVPLEPVDQTLDRCOEIPGCVGVVPGAGYDAIAYLVLENQGNFKQ--KTL 400
DB 409 MGEAAQVLEPESQIRLLDTTMMMDGVLAGVPGAGGFDAVAVTIGDSCTVAKAMSSL 468
QY 401 ENPDYFHNYYWDLBEQTEGVLEE 424
DB 469 -----NVLALLVREDPFGVLLLE 485

RESULT 9
AAB5063 standard; Protein; 368 AA.
XX AAB5063;
XX 13-FEB-2002 (first entry)
XX Enterococcus faecalis cellular proliferation protein #350.
XX DE Antisense; prokaryotic cellular proliferation protein;
XX KM antibiotic; antibacterial; drug design.
XX OS Enterococcus faecalis.
XX PN MO200170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX DR MPI: 2001-611495/70.
XX DR N-PSDB; AAS52922.
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX PS Example 3; Seq ID No 10656; 511pp; English.
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
```

```
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 368 AA;
XX
XX Query Match 9.4%; Score 209.5; DB 22; Length 368;
XX Best Local Similarity 27.0%; Pred. No. 2.1e-10;
XX Matches 113; Conservative 52; Mismatches 137; Indels 117; Gaps 23;
QY 6 SAPKRAFLAGYLVLEPIYDAVYVTLSSRMAVITPKGSLKSRIKISPPQA----- 59
DB 5 TTRGKLFAGEYAVVPEPGHPIIVAVD--QFVTVVEETDGG--IQSAQYSLPIRW 59
QY 60 ---NGEMBYHISNTEKREBVOSRINPLEATIFVLAVIQTEAPDLIITYSDPGHYS 116
DB 60 TRRNGELVLDI-----RENPF-----HYVLAAILHILTEKY-----A 89
QY 117 QEDTETKTSNCKEPTLYSRATEVE-----KTGLSSAGLYSVVATSLSHPIPNVIS 171
DB 90 QE-----QNKELSF-YHLKVTSLDSSNGKRYGLSS--GAVTVGYKALNIFYD--LG 138
QY 172 TNKDIHNAVQIARHCYAKKIGSGFVATYATY-GLIYRRFQPALINDVFOVLESDEPEK 230
DB 139 LENBEIFPLSLAHLAVQGN-GSCGDIAASCYCGWIASTPHDMVN-----QKV 187
QY 231 PTE-LKKLIENWEE-----XHERCTLPYGIKILMGDVKGSETPKLVSRVLQMKKEK 283
DB 188 TTETLTDLAMDPELMIFPLK-----VPKQLRLIGWTGSPASTSDLVDRHQSKEEK- 241
QY 284 EESSVYVQIUNANLQPKELREMKEXDSDPTYIKELDSVEPLTVAIKNIRKGLQA 343
DB 242 -----QAYEQFLMKSRLC-----VETWIN--GFVTGKISVIOKQITKRRQL 282
QY 344 TOKSE---VPIEPVDQTLDRCOEIPGCVGVVPGAGYDAIAYLVLENQGNFKOKT 399
DB 283 AELSSLTGVIETELKRLCDLABSYTGAAS--SGAGGGDGGIYI-----FRQKS 331

RESULT 10
AAB60859
ID AAB60859 standard; protein; 368 AA.
XX AAB60859;
XX 30-MAR-2001 (first entry)
XX DE Mevalonate pathway protein #3.
XX KM Mevalonate pathway; disease; infection.
XX OS Enterococcus faecalis.
XX PN MO200078935-A1.
XX PD 28-DEC-2000.
XX PF 22-JUN-2000; 2000WO-US17262.
XX
```

PR 22-JUN-1999; 99US-0140519.
XX 02-AUG-1999; 99US-0146682.
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX (SMIK) SMITHKLINE BEECHAM PLC.
PI Brown JR, Gwynn M, Mathie TB, Myers JE, Traini CM, Van Horn S;
XX Wilding EI;
DR MPI; 2001-071392/08.
PT New isolated mevalonate pathway gene polynucleotide derived from
PT bacterium is useful for treatment of bacterial infection -
XX
XX
PS Claim 20; Page 14; 158pp; English.
CC The present invention relates to an isolated mevalonate pathway gene
CC derived from a bacterium from clade of Class II of the
CC phylogenetic tree referred to in the specification. The invention may be
CC used for treatment of disease related to bacterial infection, e.g.
CC conjunctivitis, pneumonia, bacteraemia and meningitis.
XX
SO Sequence 368 AA;
Query Match 9.4%; Score 209.5; DB 22; Length 368;
Best Local Similarity 27.0%; Pred. No. 2.1e-10;
Matches 113; Conservative 52; Mismatches 137; Indels 117; Gaps 23;
QY 6 SARGKAFLLAGGYLVLEPIYAYTALSSRHAVITKGTSLKSRKISSPQRA----- 59
DB 5 TTGKLFINGEYAVVEGHPAIIIVAD--QFTVVEETDEGS---TQSAQYSSLPFRW 59
QY 60 ---NGMEVHISNTKPREVOSRINPFLAIFIVLAVIOTPEAFDLEIIIVSDPGYS 116
DB 60 TRNGELVLDI-----RENPF-----HYLLAIHLTEKY-----A 89
QY 117 QEDTEKTSNGEKTFLYHSRAITEVE---KTGLSSAGLVSVVATSLSHRIPNVIS 171
DB 90 QE-----QNKELSF-YHLKVTSELDSSNGRKYGSS-GAVTVGVKALNIFVD--IG 138
QY 172 TNDIILHNAQIHAACVAKKIGSGPDVATAIY-GLIVRRFQPALINDVQVLESDEPK 230
DB 139 LENEELFKLSALHIAVQGN-GSCGDIASCYGMIAFSTFDHDVNV-----QKV 187
QY 231 PTE-LKKLIESNME-----KHERCTLPYGIKILMGDVKGSETPKLVSRVLQWKKEKP 283
DB 188 TTETLIDLAMPBELMIFPLK-----VPKQKLLIGWTGSPASTSDVDRVQSKBEK- 241
QY 284 EESSVYVDQLSANLQFMKELREKREKYSDEPTYIKELDHSVEPLTVAIKIRKGLQAL 343
DB 242 -----QAAYEQFLMKSRLC-----VETMIN--GFNTGKISVIGKQITTKRQLL 282
QY 344 TQKSE---VPREDVQTLDRQCEIFGCVGVPPAGGVDALAVLLENQVNFQOKT 399
DB 283 AELISLTVGVITEALKNLCIDLAESYTGAKS--SGAGGDCGIVL-----FRQKS 331
RESULT 11
AAB60882
ID AAB60882 standard; protein; 358 AA.
XX
XX AAB60882;
XX
DT 30-MAR-2001 (first entry)
XX
DE Mevalonate pathway protein #5.
XX
KM Mevalonate pathway; disease; infection.
XX
OS Staphylococcus haemolyticus.
XX
XX WO200078935-A1.
XX

PD 28-DEC-2000.
XX
XX 22-JUN-2000; 2000WO-US17262.
XX
XX
PR 22-JUN-1999; 99US-0140519.
XX 02-AUG-1999; 99US-0146682.
XX (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PI Brown JR, Gwynn M, Mathie TB, Myers JE, Traini CM, Van Horn S;
XX Wilding EI;
DR MPI; 2001-071392/08.
PT New isolated mevalonate pathway gene polynucleotide derived from
PT bacterium is useful for treatment of bacterial infection -
XX
XX
PS Claim 20; Page 34; 158pp; English.
CC The present invention relates to an isolated mevalonate pathway gene
CC derived from a bacterium from clade of Class II of the
CC phylogenetic tree referred to in the specification. The invention may be
CC used for treatment of disease related to bacterial infection, e.g.
CC conjunctivitis, pneumonia, bacteraemia and meningitis.
XX
SO Sequence 358 AA;
Query Match 8.7%; Score 193; DB 22; Length 358;
Best Local Similarity 24.0%; Pred. No. 6.7e-09;
Matches 101; Conservative 63; Mismatches 129; Indels 128; Gaps 21;
QY 7 AFGKAFLLAGGYLVLEPIY-----DAYTALSSRHAVIT-----PKGTSLKES 49
DB 6 AFGKLVAGEYAVTEPGYSVLIAVDRFTASIEASNAVTSITHSKTHIYBEVTRRND 65
QY 50 RIKISSPOPANGMEVHISNTKPREVOSRINPFLAIFIVLAVIOTPEAFDLEIIIV 109
DB 66 KIDISDANASQ-----LKVVTAEVFEQYARSCNVKL----- 99
QY 110 SPQCHSGEDTETKTSNGEKTFLYHSRAITEVEKTCGSSAG-LVSVVATSLSHRIPN 168
DB 100 --KHFLLEIDSLNLDASG-----NRYGLSSAAVIVSVV--KALNEFYDM 140
QY 169 VISTNKDILHNAQIHAACVAKKIGSGPDVATAIY-GLIVRRFQPALIND-VFOVLES 226
DB 141 QLSNL--YIKLAVISNNRLOSLSCG-DIAVSIVSGMLAISTFD---HDVVKQOMER- 192
QY 227 PEKFTTELKLLIESNMEKH-ERCTLPYGIKILMGDVKGSETPKLVSRVLQWKKEKPE 285
DB 193 -----TSVNEVLEKWPGLHIEPLQAPENMEVLIGWTGSPASPHLVSEVKRLK----- 241
QY 286 SSVYVDQLSANLQFMKELREKREKYSDEPTYIKELDHS---VEPLVATK--NIRKGL 340
DB 242 -----SDPSYGRFLQSHTCVNLNLYAFETDNI-KGV 273
QY 341 QALTQKSEVPIEP-----DVQTLDRQCEIFGCVGVV--VPAGGVDALAVLLENQ 391
DB 274 QCMIRONRMIIOQMNENATVDIETENLKLCDIGERYGAATAKTSAGGDC-GIALINDR 332
QY 392 V 392
DB 333 I 333
RESULT 12
ABP53755
ID ABP53755 standard; Protein; 179 AA.
XX
XX ABP53755;
XX
DT 03-JAN-2003 (first entry)
XX
XX

DE Rice mevalonate kinase SEQ ID NO:22.
 XX Mevalonate kinase; enzyme; herbicide; agricultural; plant.
 KW Oryza sativa.
 OS US2002119546-A1.
 XX
 PN 29-AUG-2002.
 XX
 PD 20-JUL-2001; 2001US-0909745.
 PF
 XX 05-NOV-1998; 98US-107241P.
 PR 04-NOV-1999; 99US-0433242.
 XX
 PA (PALC/) PALCO S C.
 PA (FAMO/) FAMODU O O.
 XX
 PI Falco SC, Famodu OO;
 XX
 DR WPI; 2002-731357/79.
 DR N-PSDB; ABO82712.
 XX
 PT New mevalonate kinase polypeptide useful for new herbicide discovery
 PT and design, and for altering levels of mevalonate kinase in transformed
 PT cells -
 XX
 PS Example 4; Fig 3; 39pp; English.
 XX
 CC The present invention describes an isolated mevalonate kinase polypeptide
 CC (I), having a sequence identity of 80 % based on the Clustal method of
 CC alignment when compared to corn, rice, soybean and wheat mevalonate
 CC kinase. (II) has herbicide and agricultural activities. A nucleic acid
 CC (III) encoding (I) is useful for transforming a cell and producing a
 CC transgenic plant. A chimeric gene (III) comprising (II) is useful for
 CC altering the level of expression of a squalene biosynthetic enzyme in a
 CC host cell, where the expression of the chimeric gene results in
 CC production of altered levels of mevalonate kinase in the transformed host
 CC cell. (III) is also useful for evaluating a compound for its ability to
 CC inhibit the activity of a mevalonate kinase. (I) is useful as a target
 CC to facilitate the design and/or identification of inhibitors of these
 CC enzymes that may be useful as herbicides. Fragments of (II) are useful
 CC to create transgenic plants in which the disclosed polypeptides are
 CC present at higher or lower levels than normal or in cell types or
 CC developmental stages in which they are not normally found. The nucleic
 CC acid fragments are also useful as probes for genetically and physically
 CC mapping the genes that they are a portion of and as markers for traits
 CC linked to those genes. The present sequence represents a rice mevalonate
 CC kinase protein from the present invention.
 CC
 CC Sequence 179 AA;
 SQ
 Query Match 8.1%; Score 181.5; DB 23; Length 179;
 Best Local Similarity 31.5%; Pred. No. 2.6e-08;
 Matches 47; Conservative 21; Mismatches 56; Indels 25; Gaps 2;
 QY 264 GGSSTPKLVSRVYQKKEKPESSVYDQNSANLQFMKELEBEMKEDSDPEYIKELD 323
 DB 1 GGSSTPSNVGVSKQKQKSDPKSKETWKLGTIANVLENLNKLAEDHWEATESVL- 59
 QY 324 HSEVPLTV-----AIKIRKGLQALQKSEVPLEPDVOTQL 359
 DB 60 RSCSRLTCSKMTETAVTNOHQLIVRSLLAARDAFLRHLHREMGIAAGVPIEPSQQL 119
 QY 360 LDRCEIPGCVGVVPGAGVDAIVLV 388
 DB 120 LDATNMGEVLLAGVPGAGGFDAVSVLL 148
 RESULT 13
 AAG81912
 ID AAG81912 standard; Protein; 358 AA.
 XX

AC AAG81912;
 XX
 XX 03-SEP-2001 (first entry)
 XX
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:918.
 XX
 KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN WO200134809-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 09-NOV-2000; 2000WO-US30782.
 XX
 PR 09-NOV-1999; 99US-0164258.
 XX
 PA (GLAX) GLAXO GROUP LTD.
 XX
 PI kimberly WT;
 XX
 DR WPI; 2001-316495/33.
 DR N-PSDB; AAH52762.
 XX
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis -
 XX
 PS Claim 18; Page 273; 2188pp; English.
 XX
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 CC
 CC Sequence 358 AA;
 SQ
 Query Match 8.1%; Score 181.5; DB 22; Length 358;
 Best Local Similarity 24.1%; Pred. No. 7.7e-08;
 Matches 99; Conservative 52; Mismatches 131; Indels 129; Gaps 18;
 QY 7 AFGAFLAGGYLVLEPIYDAYVATLSSRMHNV-----TPKSTSLKES 49
 DB 6 AFGKLYIAGEVAYATEPGKSTILVAVNFVATIASNKVSGSHSKTLHYEPVAFDRRED 65
 QY 50 RIKISSFOFANGEMEHYHSSNTEKPREVQSKINFLFETIFVLAIVQPTAFPLEIITY 109
 DB 66 RIKISDVQAAKQ-----LKYVVAIVFQDYVSSCMNNT--- 99
 QY 110 SDPYHSQEDTETTSNGEKTFLYHSRALTEVEKTLGSSAG-LVSVVATSLSHFTPN 168
 DB 100 --KHFLHTLDSNLDNG-----QKTLGSSAAVTVSVV--KALNEFTG- 139
 QY 169 VISTNKDILHNVAQIAHCAVQKTIQSGFEDVATAIY-GLIYVRFQPALIND-VFQVLES 226
 DB 140 -LELSNLIYIKLVAVIANMKLQSLSSCG-DIAVSYSGLAVSTFD---HDMVQQMBE- 192

QY 227 PEKFTLKLIESNNEEKH-ERCTLPYGIKLMGVKGSSTPKLVSRYLQMKKEKPEE 285
DB 193 -----TSVNDVLEKMPGHLIEPLQAPENNEVLIGWTGSPASSPHLVSFKRLK----- 241
QY 286 SSVVYDQNSANLQFMKELREMERKDYSDPEYIKELDHS---VEPLTVAIK--NI----- 336
DB 242 -----SDPSFYGDFLDQSHACVESLIQAFKTNINIKGVQ 274
QY 337 -----RKGLQALTKQSEVPPIEPDVOTQLDRCOEIPGCVGVVPGAGGYD 381
DB 275 KMIIRNRRIIIGQMDNEASVIEITDKKLCVGEKRGK--ASKTSGAGGD 323

RESULT 14
AAB60883
ID AAB60883 standard; protein; 358 AA.
XX
AC AAB60883;
XX
DT 30-MAR-2001 (first entry)
XX
DE Mevalonate pathway protein #5.
XX
KM Mevalonate pathway; disease; infection.
XX
OS Staphylococcus epidermidis.
XX
PN MO200078935-A1.
XX
PD 28-DEC-2000.
XX
PF 22-JUN-2000; 2000MO-US17262.
XX
PR 22-JUN-1999; 99US-0140519.
XX
PR 02-AUG-1999; 99US-0146682.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Brown JR, Gwynn M, Mathie TB, Myers JE, Traini CM, Van Horn S;
PI Wilding EI;
XX
DR WPI; 2001-071392/08.
XX
PT New isolated mevalonate pathway gene polynucleotide derived from
PT bacterium is useful for treatment of bacterial infection -
XX
XX
PS Claim 20; Page 34-35; 158pp; English.
XX
CC The present invention relates to an isolated mevalonate pathway gene
CC derived from a bacterium from clade of Class II of the
CC phylogenetic tree referred to in the specification. The invention may be
CC used for treatment of disease related to bacterial infection, e.g.
CC conjunctivitis, pneumonia, bacteraemia and meningitis.
XX
SQ Sequence 358 AA;

Query Match 8.1%; Score 181.5; DB 22; Length 358;
Best Local Similarity 24.1%; Pred. No. 7.7e-08;
Matches 99; Conservative 52; Mismatches 131; Indels 129; Gaps 18;

QY 7 APGKAPLAGGYLVLEPIYDAYVTALSRMHAVI-----TPKGTSLKES 49
DB 6 APGKLYIAGEYAVTEBGYSILAVNRFTVATIEASNKVGSIHSTKLTAHEPVKPRNED 65
QY 50 RIKISSPQFANGMEWHYISSNTEKPREVOSRINPFLERATIFVLAYIQPTAEADLEIITY 109
DB 66 RIEISVQAAKQ-----LKVVTAIEVEPQYVRSQNMML--- 99
QY 110 SDGCHYSQETERTKTSNGEKTLVYHSRAATEVEKGLGSSAG-LYSVATSLISHFIPN 168
DB 100 --KHFIITLDSNLADNSG-----QKYGLAGSSAAVLVSVV--KALINERYG- 139

QY 169 VISTNKDILAHNAQIAHCAQKIGSGPDVATAY-GLIVYRFPQALIND-VFOVLESD 226
DB 140 -LELSNLYIKLAVIANMQLQSLSSGQ-DIANSVSGMLAYSTPD---HDWVKQMEB- 192
QY 227 PEKFTLKLIESNNEEKH-ERCTLPYGIKLMGVKGSSTPKLVSRYLQMKKEKPEE 285
DB 193 -----TSVNDVLEKMPGHLIEPLQAPENNEVLIGWTGSPASSPHLVSFKRLK----- 241
QY 286 SSVVYDQNSANLQFMKELREMERKDYSDPEYIKELDHS---VEPLTVAIK--NI----- 336
DB 242 -----SDPSFYGDFLDQSHACVESLIQAFKTNINIKGVQ 274
QY 337 -----RKGLQALTKQSEVPPIEPDVOTQLDRCOEIPGCVGVVPGAGGYD 381
DB 275 KMIIRNRRIIIGQMDNEASVIEITDKKLCVGEKRGK--ASKTSGAGGD 323

RESULT 15
ABP40510
ID ABP40510 standard; Protein; 362 AA.
XX
AC ABP40510;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:535.
XX
KM Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX
KW antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-055779P.
XX
PR 08-NOV-1997; 97US-064964P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2002-381255/41.
XX
DR N-PSDB; ABN93055.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX
PS Disclosure; SEQ ID 5355; 267pp; English.
XX
XX
CC ABN90538 to ABN93174 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP3124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 362 AA;

Query Match 8.1%; Score 181.5; DB 23; Length 362;
Best Local Similarity 24.1%; Pred. No. 7.8e-08;
Matches 99; Conservative 52; Mismatches 131; Indels 129; Gaps 18;

QY 7 APGKAPLAGGYLVLEPIYDAYVTALSRMHAVI-----TPKGTSLKES 49

XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 345 AA:

Query Match 7.8%; Score 173; DB 22; Length 345;
 Best Local Similarity 22.8%; Pred. No. 4.4e-07;
 Matches 97; Conservative 57; Mismatches 143; Indels 128; Gaps 20;

QY 7 APGKALAGGYLVLEPIYDAYVTALSSRMNAVITP-----KGT-----SLKES 49
 DB 3 APGKLYAGEYAVTPEGYKSVLIALDRFYATIEETDQYKGIHSAKALHNPVTFSRBD 62
 QY 50 RIKISSPOFANGEMWEHISNTEKPREVOSRINPLEATIFIVLAIOTPEAFDEIITY 109
 DB 63 SIVISDPHAKQ-----LNVVTAIEIFEQYAKSCDIAM---96
 QY 110 SDPGYHSQEDTETKSSNGEKTFLYHSRAITEVEKTGLSSAG-LVSVVATSLSHPIFN 168
 DB 97 --KHFLITIDSNL--DSSNG-----HKYGLSSAAVLVSVI--KVINEFYDM 137
 QY 169 VISTNDILHNAVAQIAHCYAOKKIGSGPDVATAY-GLIVRRFOPALINDVQVLESBP 227
 DB 138 KLSNL--YIYKLAIVANMKLQSLSSCG-DIAVSYSGLAVSTFDEHWKQIE-----188
 QY 228 EKPEPELAKLIESNMEKH-ERCTLPYGIKLMGDVKGSETPKLVSRVLOWKKEPES 286
 DB 189 ---DITVEEVLINWGCHIEPIQAPENNEVLIGTWGSPASSPHFSEVKRLK-----238
 QY 287 SVVYDQLSANLQFMKELREMERREKYSDEPTY--IKELDSVPEPLTVAIK-NIRKGLQA 342
 DB 239 -----SDPSFYGDPLEDSHRCVEKLIHAFKTNINIKGVOK 272
 QY 343 LNOKSEVPTE-----PVQTO---LIDRCOEIGCGVGVPRAGCDALAVLLENO 391
 DB 273 MVRQNTITIQMDKEATVIEIEKTKYLCIDIAEKYHG--ASKTSAGAGDQC-GITLINRO 329
 QY 392 VGNFK 396
 DB 330 VDKEK 334

RESULT 18
 AAU37273
 ID AAU37273 standard; Protein; 358 AA.

AC AAU37273;

DE 14-FEB-2002 (first entry)

XX Staphylococcus aureus cellular proliferation protein #1443.

XX Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

XX

OS Staphylococcus aureus.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

XX 26-MAY-2000; 2000US-207727P.

XX 23-OCT-2000; 2000US-242578P.

XX 27-NOV-2000; 2000US-253625P.

XX 22-DEC-2000; 2000US-257931P.

XX 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,
 PI Yamamoto RT, Xu HH;
 XX WPI; 2001-611495/70.
 DR N-PSDB; AASS5132.

PT New polynucleotides for the identification and development of
 XX antibiotics, comprise sequences of antisense nucleic acids -

PS Example 3; Seq ID No 12866; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 358 AA:

Query Match 7.8%; Score 173; DB 22; Length 358;
 Best Local Similarity 22.8%; Pred. No. 4.4e-07;
 Matches 97; Conservative 57; Mismatches 143; Indels 128; Gaps 20;

QY 7 APGKALAGGYLVLEPIYDAYVTALSSRMNAVITP-----KGT-----SLKES 49
 DB 6 APGKLYAGEYAVTPEGYKSVLIALDRFYATIEETDQYKGIHSAKALHNPVTFSRBD 65
 QY 50 RIKISSPOFANGEMWEHISNTEKPREVOSRINPLEATIFIVLAIOTPEAFDEIITY 109
 DB 66 SIVISDPHAKQ-----LNVVTAIEIFEQYAKSCDIAM---99
 QY 110 SDPGYHSQEDTETKSSNGEKTFLYHSRAITEVEKTGLSSAG-LVSVVATSLSHPIFN 168
 DB 100 --KHFLITIDSNL--DSSNG-----HKYGLSSAAVLVSVI--KVINEFYDM 140
 QY 169 VISTNDILHNAVAQIAHCYAOKKIGSGPDVATAY-GLIVRRFOPALINDVQVLESBP 227
 DB 141 KLSNL--YIYKLAIVANMKLQSLSSCG-DIAVSYSGLAVSTFDEHWKQIE-----191
 QY 228 EKPEPELAKLIESNMEKH-ERCTLPYGIKLMGDVKGSETPKLVSRVLOWKKEPES 286

Db 192 ---DTTVEEVLKNNPGLHIEPLQAEENNEVLIGWTSGPASSPHFVSEVKRLK----- 241
QY 287 SVVYDQNSANLQFMKELEREMEKYSDEPTY---IKELDHSVEPLTVAIK-NIRKGLQA 342
Db 242 -----SDPSFYGDPLEDSHRCVEKLIHAFKTNNIKGVOK 275
QY 343 LTQKSEVPLE-----PDVQTQ---LLDRQCEIPGCVGVPPAGGYDAIAVLVLNQ 391
Db 276 MVRQNRRTIIRQMDKEATVDIETEKLYCDIAEKYHG--ASKTSGAGGDC-GITITIND 332
QY 392 VGNFK 396
Db 333 VDKK 337

RESULT 19
AAB60864
ID AAB60864 standard, protein, 358 AA.
XX AAB60864;
AC
XX
XX
DT 30-MAR-2001 (first entry)
XX
DE Mevalonate pathway protein #2.
XX
XX Mevalonate pathway; disease; infection.
XX OS Staphylococcus aureus.
XX PN W0200078935-A1.
XX PD 28-DEC-2000.
XX PF 22-JUN-2000; 2000WO-US17262.
XX PR 22-JUN-1999; 99US-0140519.
XX PR 02-AUG-1999; 99US-0146882.
XX PA (SMIK) SMITHKLINE BEECHAM CORP.
XX PA (SMIK) SMITHKLINE BEECHAM PLC.
XX PI Brown JR, Gwynn M, Mathie TB, Myers JE, Traini CM, Van Horn S;
XX PI Wilding EI;
XX DR WPI; 2001-071392/08.
PT New isolated mevalonate pathway gene polynucleotide derived from
XX bacterium is useful for treatment of bacterial infection -
XX
XX PS Claim 20; Page 17-18; 158pp; English.
XX CC The present invention relates to an isolated mevalonate pathway gene
XX CC derived from a bacterium from clade II of the
XX CC phylogenetic tree referred to in the specification. The invention may be
XX CC used for treatment of disease related to bacterial infection, e.g.
XX CC conjunctivitis, pneumonia, bacteremia and meningitis.
XX SQ Sequence 358 AA;

Query Match 7.8%; Score 173; DB 22; Length 358;
Best Local Similarity 22.8%; Pred. No. 4,6e-07;
Matches 97; Conservative 57; Mismatches 143; Indels 128; Gaps 20;

7 AAGKAFIAGGYLVLEPIYDAYVTALSRMHAYITP-----KGT-----SLAES 49
Db 6 AAGKAYIAGEYAVTBPGKSVIALIDRFVTATIEBATQYKGIHSAKALHNPVTFSRDED 65
QY 50 RIKISSPPFANGEMWEYHISSTNEKREYOSRINPFLAETIFVLAVIQPTFAFLEITITY 109
Db 66 SIVISDPAHAKO-----LNVVYVAIEIFEEGYAASCDIAM--- 99
QY 110 SDPGHSHQEDTETKSSNGEKTFLYHSRAITEVEKTGLASSAG-LVSVVAISLSLHPIPN 168

Db 100 --KHFHLITDSNL-DDSNG-----HKYGLSSAAVAVSVL--KVINEFYDM 140
QY 169 VISTNKDILHNVAOIAHCAOKKIGSPDVATAIY-GLIVRRFPQALINDVFOYLESDP 227
Db 141 KLSNL--YIKLAVIANNKLOSLSCG--DIAVSISGWLAVSTFHEWYKQIE----- 191
QY 228 EKFPTELKRLIESNNEKH-ERCTLPYGIKILMGDVKGSETPKLVSRYLQWKEKEPERS 286
Db 192 ---DTTVEEVLKNNPGLHIEPLQAEENNEVLIGWTSGPASSPHFVSEVKRLK----- 241
QY 287 SVVYDQNSANLQFMKELEREMEKYSDEPTY---IKELDHSVEPLTVAIK-NIRKGLQA 342
Db 242 -----SDPSFYGDPLEDSHRCVEKLIHAFKTNNIKGVOK 275
QY 343 LTQKSEVPLE-----PDVQTQ---LLDRQCEIPGCVGVPPAGGYDAIAVLVLNQ 391
Db 276 MVRQNRRTIIRQMDKEATVDIETEKLYCDIAEKYHG--ASKTSGAGGDC-GITITIND 332
QY 392 VGNFK 396
Db 333 VDKK 337

RESULT 20
ABB49817
ID ABB49817 standard, protein, 359 AA.
XX ABB49817;
AC
XX
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes protein #2521.
XX
XX
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX KM Vitamin B12; bacterial infection; disease.
XX OS Listeria monocytogenes.
XX PN W0200177335-A2.
XX PD 18-OCT-2001.
XX PF 11-APR-2001; 2001WO-FR01118.
XX PR 11-APR-2000; 2000FR-0004629.
XX PA (INSP) INST PASTEUR.
XX PI Buchrieser C, Frangeul L, Couve B, Rusniok C, Fajhi H, Dehoux P,
XX PI Duseurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P,
XX PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
XX PI Dominguez-Bernal G, Garrido-Garcia P, Tietres-Martinez A, Amend A,
XX PI Charabarty T, Dommann E, Hain T, Berche P, Charbit A, Durant L,
XX PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,
XX PI Mademonte E, De Pablo B, Wehlend J, Kaerst U, Entian K, Hauf J;
XX PI Rose M, Voss H;
XX DR WPI; 2002-010914/01.

Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX and prevention of Listeria and related bacterial infections, and
XX related polypeptides -
XX
XX PS Claim 6; SEQ ID No 2522; 192pp; French.

The present invention relates to the genome sequence of Listeria
XX monocytogenes EGD-e (see ABA01041). The genome sequence and fragments of
XX it are useful for selecting probes and primers for detecting genes in L.
XX monocytogenes and related organisms, and for studying genetic
XX polymorphisms and other genomes. The present sequence is a protein
XX encoded by the genome sequence of the present invention. Proteins
XX expressed from the genome sequence are useful for raising specific
XX antibodies, identification of L. monocytogenes and related organisms, and

CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by L.
 CC monocytogenes and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 359 AA;

Query Match 7.6%; Score 170.5; DB 23; Length 359;
 Best Local Similarity 24.7%; Pred. No. 7.9e-07;
 Matches 96; Conservative 55; Mismatches 144; Indels 93; Gaps 19;

QY 8 PGKAFIAGGYLVLEPIYDAYVTALSSRMHAVITPKGSLKSRKIKISSPOFANGEWYHI 67
 DB 13 PGKLVYAGEYAVVESGHTALTITA---VNRVIT---LTLEDSE-----RHELMIPHY 57
 QY 68 SSTXTEKREYQSRINP-----PLEATIFIVAVIQTEAFDL---EIIYSDPGYHSQE 118
 DB 58 ENFVSMP--VGGELKPDGEHMTTAAINATTFK--SEGIELTPVKMI-----104
 QY 119 DTEKTSNNGEKFVLYSRATIEVEKTKGLSSAGLVSVVATSLSHPIPNVISTNKDILH 178
 DB 105 EFTLIDIOSGA-----KYGSSAAATVAVVAINALMTFTYPISTLTKR---F 146
 QY 179 NVAQIAHCYAAOKKIGSGPDVATAY-GLIVRRFPALINDVFOVLESDEKPEPTELK 237
 DB 147 KLAALSHLVYQGN-GSCGDIAACMGGMIAVTFPDGEWK--HRLAVKSLSEWFKSPWPM 203
 QY 238 IESNMEEKHRCRLPYGIKLMGVKGSSETP---KLVSRYQMKKEKEBESSVYDOL 293
 DB 204 LQI---ETLEBPVPTFSV-----GWTGTPVSTGSLVSOIHAFK---QEDSKRYQHF 248
 QY 294 NSNLTQFMKELREMRREKYDSDPEYTIKELDHVSVEPLTVAIKNIRKGLQALTKSEVPIEP 353
 DB 249 LTFNNEIMQOITQAFHTKDE-----ELYSATIKENRRILQELGTAGVNIET 295
 QY 354 DVQTOILLDRQCEIPGCVGVVPGAGYD 381
 DB 296 SLIKELADSAENMG--AGKSGSGGCD 321

RESULT 21

ID AAB99728 standard; Protein; 374 AA.

XX AAB99728;
 AC XX
 DT 10-SEP-2001 (first entry)
 XX XX
 DE Streptomyces sp. CL190 mevalonate pathway orfc protein SEQ ID NO:10.
 XX XX
 KM Streptomyces sp. CL190: mevalonate pathway; actinomycete; cardiant;
 KM isoprenoid compound; osteopathic; cytostatic; ubiquitome; vitamin K2;
 KM carotenoid; heart disease; osteoporosis; cancer; drug; health food..
 OS Streptomyces sp. CL190.
 XX XX
 PN WO200142476-A1.
 XX XX
 PD 14-JUN-2001.
 XX XX
 PF 06-DEC-2000; 2000WO-JP08620.
 XX XX
 PR 08-DEC-1999; 99JP-0348375.
 XX XX
 PA (SETO/) SETO H.
 PA (KUZU/) KUZUYAMA T.
 XX XX

PI Seto H, Kuzuyama T, Takahashi S, Takagi M;
 XX XX
 DR MPI: 2001-381696/40.
 DR N-PSDB; AAH44046.
 XX XX
 PT Actinomycetes-originated genes of enzymes participating in mevalonate
 PT pathway, applicable in producing e.g. ubiquitome, vitamin K2 and
 PT carotenoids for treatment of heart diseases, osteoporosis and cancer in
 PT drug and health food -
 XX XX
 PS Claim 6; Page 62-64; 75pp; Japanese.

CC The sequence given in AAH44043 represents a DNA sequence isolated from
 CC Streptomyces sp. CL190, containing a 6798 base pairs (S1), which encodes
 CC the whole enzyme necessary for functioning the mevalonate pathway. The
 CC sequence encodes protein sequences, designated orfc to E and hmgir, which
 CC are used in the mevalonate pathway. The proteins and polynucleotide
 CC sequences encoding them have cardiant, osteopathic and cytostatic
 CC activities. The genes are applicable in producing e.g. ubiquitome,
 CC vitamin K2 and carotenoids which can be used in the treatment of heart
 CC diseases, osteoporosis and cancer in drugs and health foods. The present
 CC sequence represents the orfc protein from the present invention.

XX SQ Sequence 374 AA;

Query Match 6.6%; Score 147; DB 22; Length 374;
 Best Local Similarity 22.4%; Pred. No. 0.00012;
 Matches 100; Conservative 61; Mismatches 161; Indels 124; Gaps 23;

QY 7 APKAFIAGGYLVLEPIYDAYVTALSSRMHAVITPKGSLKSRKIKISS---POFANGEW 63
 DB 12 APKGLFVAGEYAVVDPGNPAIIVAVDRHISVTSADADTGAADVITSDDLQPAVG--W 69
 QY 64 EHH---ISSNTEKREYQSRINPFLSATIFIVAVIQTEAFDEIIYSDPGYHSQED 119
 DB 70 RWHDRLLVVRDPPDDQQAASALAHVVSAL-----ET 100
 QY 120 TETKTSNNGEKFVLYSRATIEVEKTKGLSSAGLVSVVATSLSHPIPNVISTNKD 175
 DB 101 VGRLLGERQKYPALITLSSSRLLHEDGRKFGSS--GAVTVATVAANAFCLEIST--D 157
 QY 176 IAHNVAQIAHCYAAOKKIGSGPDVATAY-GLIVRRFPALINDVFOVLESDEKPEPTEL 234
 DB 158 ERFRLAMLATAEIDPK-GSGGDLAASWGMIAVQAPRAFAVLDLARRGV-----208
 QY 235 KGLISNNE-EKHERCTLPYGIKLMGVKGSSETPKVSRYLQ--WKKEKEBESSV--V 289
 DB 209 -RTLKAPWPGHSVRRLPAKGLTLEVGTGEPASTASLVSDLRRTWGSASHORFVETT 267
 QY 290 YDOLNSA-----NLQFMKELREMRREKYDSDPEYTIKELDHVSVEPLTVAIKNIRKGLQ 341
 DB 268 TDCVRSAYVALESQDTSIHRIIRAROE-----LARLDDV-----GLG 307
 QY 342 ALTKSEVPEIPDVOTQLDRQCEIPGCVGVV--PGAAGYD-AIAYVLB--NOVGNFK 396
 DB 308 IFTPK-----LTALCDAAE---AVGGAAPSGAGGCGIALDLDAEASRDITVHR 354
 QY 397 QKTLENPDYFNHVVVWVDEEQTEGVL 422
 DB 355 QR-----W-----ETAGVL 363

RESULT 22

ID AAB18130 standard; Protein; 109 AA.

XX AAB18130;
 AC XX
 DT 08-NOV-2000 (first entry)
 XX XX
 DE Pinus radiata phosphomevalonate kinase protein SEQ ID NO:291.
 XX XX
 KW Eucalyptus grandis; Pinus radiata; modification; isoprenoid; plant;

KW metabolism; isoprenoid biosynthetic pathway; terpenoid; steroid;
 KW genome mapping; physical mapping; positional cloning; forestry;
 KW agriculture; medicine; fermentation; plant development; pest resistance;
 KW pinene; mycene; Monterey pine.
 XX
 OS Pinus radiata.
 XX
 PN WO200036081-A2.
 XX
 PD 22-JUN-2000.
 XX
 PF 16-DEC-1999; 99WO-NZ00219.
 XX
 PR 17-DEC-1998; 98US-0215504.
 XX
 PR 29-JUL-1999; 99US-0146441.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Havukala IJ;
 XX
 DR WPI; 2000-431575/37.
 DR N-PSDB; AAA69566.
 XX
 PT New plant polynucleotides encoding polypeptides involved in the
 PT production and modification of isoprenoids, useful in forestry and
 PT agriculture for manipulation of isoprenoid metabolism -
 XX
 PS Claim 26; Page 159; 164pp; English.
 XX
 CC The present invention describes plant polynucleotides encoding
 CC polypeptides involved in the production and modification of isoprenoids,
 CC such as terpenoid and steroid compounds. The polynucleotides are used
 CC in genome mapping, in physical mapping and in positional cloning of
 CC genes. The polynucleotides and polypeptides are useful in forestry and
 CC agriculture for manipulation of isoprenoid metabolism, in medicine for
 CC therapeutic effects, including direct application in diseased organisms
 CC or indirect application by transgenic organisms and in fermentation and
 CC chemical processing industries involving isoprenoids. In plant
 CC applications, manipulating isoprenoid pathways or isoprenoid composition
 CC may, for example, affect plant development, pest resistance, and the
 CC value of extractives (e.g. pinene and mycene). The ubiquitous and
 CC varied roles of isoprenoids make the polynucleotides attractive targets
 CC for biotechnical applications in a variety of fields. AAA69527 to
 CC AAA69690 and AAA18004 to AAA18143 represent Eucalyptus grandis and Pinus
 CC radiata polynucleotides and proteins used in the exemplification of the
 CC present invention.
 XX
 XX Sequence 109 AA;
 SQ
 Query Match 5.6%; Score 124.5; DB 21; Length 109;
 Best Local Similarity 33.0%; Pred. No. 0.0021;
 Matches 37; Conservative 17; Mismatches 45; Indels 13; Gaps 4;
 QY 1 MSKAFSAPGKAFIAGCYLVLEPIYDAYVTALSSRMNAVITPKGTSIKESR-----IXI 53
 DB 1 MAVVVSAPGKVLITAGVILTEKPNGLVLTARFAIVKPRITSDSSWAMLTDTVL 60
 QY 54 SSPQFANGMEWHISSNTEKREYV--QSRNPFLEATITFYLAIOPEAF 102
 DB 61 TSPQIAK-EATYKLSIKTLSTLONVASSSSSNENPFVEQAVQFAVAAK-EAF 109
 RESULT 23
 ID AAG12901 standard; Protein; 378 AA.
 XX
 XX AAG12901;
 AC
 XX
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 12193.
 XX

KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 XX
 PR 05-MAR-1999; 99US-0123180.
 XX
 PR 09-MAR-1999; 99US-0123548.
 XX
 PR 23-MAR-1999; 99US-0125788.
 XX
 PR 29-MAR-1999; 99US-0126264.
 XX
 PR 01-APR-1999; 99US-0126785.
 XX
 PR 06-APR-1999; 99US-0127462.
 XX
 PR 08-APR-1999; 99US-0128234.
 XX
 PR 16-APR-1999; 99US-0129845.
 XX
 PR 19-APR-1999; 99US-0130077.
 XX
 PR 21-APR-1999; 99US-0130449.
 XX
 PR 23-APR-1999; 99US-0130510.
 XX
 PR 28-APR-1999; 99US-0130891.
 XX
 PR 30-APR-1999; 99US-0131449.
 XX
 PR 04-MAY-1999; 99US-0132407.
 XX
 PR 05-MAY-1999; 99US-0132484.
 XX
 PR 06-MAY-1999; 99US-0132486.
 XX
 PR 07-MAY-1999; 99US-0132487.
 XX
 PR 11-MAY-1999; 99US-0132863.
 XX
 PR 14-MAY-1999; 99US-0134218.
 XX
 PR 14-MAY-1999; 99US-0134219.
 XX
 PR 14-MAY-1999; 99US-0134370.
 XX
 PR 18-MAY-1999; 99US-0134768.
 XX
 PR 19-MAY-1999; 99US-0134941.
 XX
 PR 20-MAY-1999; 99US-0135124.
 XX
 PR 21-MAY-1999; 99US-0135353.
 XX
 PR 24-MAY-1999; 99US-0135629.
 XX
 PR 25-MAY-1999; 99US-0136021.
 XX
 PR 27-MAY-1999; 99US-0136392.
 XX
 PR 28-MAY-1999; 99US-0136782.
 XX
 PR 01-JUN-1999; 99US-0137222.
 XX
 PR 03-JUN-1999; 99US-0137528.
 XX
 PR 04-JUN-1999; 99US-0137502.
 XX
 PR 07-JUN-1999; 99US-0137724.
 XX
 PR 08-JUN-1999; 99US-0138094.
 XX
 PR 10-JUN-1999; 99US-0138540.
 XX
 PR 10-JUN-1999; 99US-0138847.
 XX
 PR 14-JUN-1999; 99US-0139119.
 XX
 PR 16-JUN-1999; 99US-0139452.
 XX
 PR 17-JUN-1999; 99US-0139453.
 XX
 PR 18-JUN-1999; 99US-0139454.
 XX
 PR 18-JUN-1999; 99US-0139455.
 XX
 PR 18-JUN-1999; 99US-0139456.
 XX
 PR 18-JUN-1999; 99US-0139457.
 XX
 PR 18-JUN-1999; 99US-0139458.
 XX
 PR 18-JUN-1999; 99US-0139459.
 XX
 PR 18-JUN-1999; 99US-0139460.
 XX
 PR 18-JUN-1999; 99US-0139461.
 XX
 PR 18-JUN-1999; 99US-0139462.
 XX
 PR 18-JUN-1999; 99US-0139463.
 XX
 PR 18-JUN-1999; 99US-0139750.
 XX
 PR 18-JUN-1999; 99US-0139763.
 XX
 PR 21-JUN-1999; 99US-0139817.
 XX
 PR 22-JUN-1999; 99US-0139899.
 XX
 PR 23-JUN-1999; 99US-0140353.

Db 240 GENTKALVSGVGRVRAHEDPAKSVFNANVDISISKEIAIITQSDFTSVTEKERIKELME 299
 Qy 325 SVEPLTVAIKIRKQALTOXSEVPIEDPVOTOLDRCOEIIPGCVGVPPAGGYDATA 384
 Db 300 MNQGLLSGVSHSISIEAV-----LITVTKHKLVSK-----LITAGGGGCVL 341
 Qy 385 VLVLENQVGNFKOKTLE 401
 Db 342 TLL---PTGTVDKVE 355

RESULT 24
 AAG50090
 ID AAG50090 standard; Protein; 378 AA.
 AC AAG50090;
 XX
 DT 18-OCT-2000 (first entry)
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 63438.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KM hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139452.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 24-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145086.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145917.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147199.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151085.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.5%; Score 123; DB 21; Length 378;
Best Local Similarity 20.1%; Pred. No. 0.02; Indels 130; Gaps 17;
Matches 88; Conservative 74; Mismatches 145; Indels 130; Gaps 17;
QY 7 APGKALAGGYLVLEPIYAVYVTLASRMHVAIT---PKGTSIKESRIKISSPQFANGEW 63

Db 7 APGKIIAGEHAV---VHGSTAVAAIDLYTVTLRFPPLPSAENNDRLLTL---QLKDIDL 60
QY 64 EYHISNTEKPREVGSRINPFLERATIFVLAIOPTREAPDLIIISDGHYSOE----- 118
Db 61 EFSWS-----LARIKEAIPYDSSTLCRSTPASCSEETLKI 96
QY 119 -----DTEKTSNGEKTFLY-HSRPI-----TEVE-KTGLSSAGIVSV 157
Db 97 AVLVERQNLPKRMMLSSGISFTFLWYTRIIIGFNPAIVYNELPYGSLSSALCVALL 156
QY 158 ATSLLSHFI-----DNVISTNKDIIHNAQIAHCAQKII---GSGFDVATAIYG 204
Db 157 TAAALASSISEKTRGNGMSLDETVELLNKVA-----FECKRIIHGKPSGIDNTWSAVG 211
QY 205 LLYRRFPQPALINDPVQVLESDEKPEPTLKKLISNMEKBERCLTPGICILMDVDVG 264
Db 212 NMI-----KFGSGEITRLSN-----MP--LMLLTNTXV 239
QY 265 GSETPKLVGRVLQWKKERPESSVYVDOLNSANLOFMKELRMRREKYDSDPTYIKELDH 324
Db 240 GNTKALVSGVQRAVRHPDAMKSVFNNAVDISIKELAAIIOKDETSVTEKERIKELME 299
QY 325 SVEPLTVAIKNIRKGLQALTQKSEVPIEDVQTOILLDRQCEIIGCVGVPPAGGYDAA 384
Db 300 MNQGLLSWGVSHSISIAV-----LITTVGHKLVSK-----LTGAGGGGCVL 341
QY 385 VLVLENQVGNFKOKTLE 401
Db 342 TLL---PTGTVDVXVVE 355

RESULT 25

ID AAGS1424 standard; Protein; 378 AA.

AC AAGS1424;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 65264.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 03-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147033.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153756.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.

```

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 5.5%; Score 123; DB 21; Length 378;
 Best Local Similarity 20.1%; Pred. No. 0.02; Indels 130; Gaps 17;
 Matches 88; Conservative 74; Mismatches 145;

```

QY 7 APGKAFAGGYLVLEPIDAYVTALSRMHAIVT---PKGSLKESRIKISSPOFANGEW 63
DB 7 APGKIIILAGEHAV---VHGSTAVAAAIIDYTYTILRFLPFAENNDRLTL---QKDIISL 60
QY 64 EYHISNTEKPREVOSRINPFLFATIFVLAYIQTEAPDLEIIYSDPGYHSQ-----118
DB 61 EFSWS-----LARIKKAIPYDSSITLCRSTPASCSEBTLKSI 96
QY 119 -----DTEKTSNGEKTELY-HSRAT-----TEVE-KTGIGSSAGLVSVV 157
DB 97 AVLVEBQNLPEKEMWMLSGISTFLMWYTRIIGFNPAVTVINSHELPGSGIGSSALCVALL 156
QY 158 ATSLSHFI-----PNVISTNKDILHNVAQIAHCYAQKRI-----GSGEPVATAIYG 204
DB 157 TTAALLASSISEKTRNGWMSLDETINBLNKMA---FEGEKTIHGKPSGIDNTVSAYG 211
QY 205 LIYVRRFOPALINDVFOVLESDEPEKPTTELKLIESNWEKHERCTLPGYGIKLMGDVYG 264
DB 212 NMI-----KFCSGEITRLQSN-----MP--LRLMLITNTRV 239
QY 265 GSETPLVSRVLQWKKEKPESSVYVDQNSANLQFMKELRMRKRYSDPEYTIKELDH 324
DB 240 GRNTKALVSGVQRAVRHPDAMKSVFNADVSIKELAAIIOKDETSVTEKERIKELME 299
QY 325 SVEPPLVIAIKNIKRGLOALTQKSEVPLEPDPVOTOLDRQCEIRPGCVGVPAGGYDATA 384
DB 300 MNQGLILSGVSHSIEAV-----LITTVKHKLVSK-----LTGAGGGGCVL 341
QY 385 VLVLENQVGNFKOKTLE 401
DB 342 TLL---PTGTVDVKVVE 355

```

```

RESULT 26
ABP53756
ID ABP53756 standard; Protein; 378 AA.
XX
AC ABP53756;
XX
DT 03-JAN-2003 (first entry)
DE Arabidopsis thaliana mevalonate kinase SEQ ID NO:23.
XX
KW Mevalonate kinase; enzyme; herbicide; agricultural; plant.
XX
OS Arabidopsis thaliana.
XX
PN US2002119546-A1.
XX
PD 29-AUG-2002.
XX
PF 20-JUL-2001; 2001US-0909745.
XX

```

```

PR 05-NOV-1998; 98US-107241P.
PR 04-NOV-1999; 99US-0433242.
XX
PA (FALC/) FALCO S.C.
PA (FAMO/) FAMODU O.O.
XX
PI Falco SC, Famodu OO;
XX
DR MPI; 2002-731357/79.

```

New mevalonate kinase polypeptide useful for new herbicide discovery
 and design, and for altering levels of mevalonate kinase in transformed
 cells -

Example 3; Fig 2; 39pp; English.

The present invention describes an isolated mevalonate kinase polypeptide
 (I), having a sequence identity of 80 % based on the Clustal method of
 alignment when compared to corn, rice, soybean and wheat mevalonate
 kinase. (II) has herbicide and agricultural activities. A nucleic acid
 (III) encoding (I) is useful for transforming a cell and producing a
 transgenic plant. A chimeric gene (III) comprising (II) is useful for
 altering the level of expression of a squalene biosynthetic enzyme in a
 host cell, where the expression of the chimeric gene results in
 production of altered levels of mevalonate kinase in the transformed host
 cell. (III) is also useful for evaluating a compound for its ability to
 inhibit the activity of a mevalonate kinase. (I) is useful as a target
 to facilitate the design and/or identification of inhibitors of these
 enzymes that may be useful as herbicides. Fragments of (II) are useful
 to create transgenic plants in which the disclosed polypeptides are
 present at higher or lower levels than normal or in cell types or
 developmental stages in which they are not normally found. The nucleic
 acid fragments are also useful as probes for genetically and physically
 mapping the genes that they are a portion of and as markers for traits
 linked to those genes. The present sequence represents an Arabidopsis
 thaliana mevalonate kinase protein from the present invention.

SO Sequence 378 AA;

Query Match 5.5%; Score 123; DB 23; Length 378;
 Best Local Similarity 20.1%; Pred. No. 0.02; Indels 130; Gaps 17;
 Matches 88; Conservative 74; Mismatches 145;

```

QY 7 APGKAFAGGYLVLEPIDAYVTALSRMHAIVT---PKGSLKESRIKISSPOFANGEW 63
DB 7 APGKIIILAGEHAV---VHGSTAVAAAIIDYTYTILRFLPFAENNDRLTL---QKDIISL 60
QY 64 EYHISNTEKPREVOSRINPFLFATIFVLAYIQTEAPDLEIIYSDPGYHSQ-----118
DB 61 EFSWS-----LARIKKAIPYDSSITLCRSTPASCSEBTLKSI 96
QY 119 -----DTEKTSNGEKTELY-HSRAT-----TEVE-KTGIGSSAGLVSVV 157
DB 97 AVLVEBQNLPEKEMWMLSGISTFLMWYTRIIGFNPAVTVINSHELPGSGIGSSALCVALL 156
QY 158 ATSLSHFI-----PNVISTNKDILHNVAQIAHCYAQKRI-----GSGEPVATAIYG 204
DB 157 TTAALLASSISEKTRNGWMSLDETINBLNKMA---FEGEKTIHGKPSGIDNTVSAYG 211
QY 205 LIYVRRFOPALINDVFOVLESDEPEKPTTELKLIESNWEKHERCTLPGYGIKLMGDVYG 264
DB 212 NMI-----KFCSGEITRLQSN-----MP--LRLMLITNTRV 239
QY 265 GSETPLVSRVLQWKKEKPESSVYVDQNSANLQFMKELRMRKRYSDPEYTIKELDH 324
DB 240 GRNTKALVSGVQRAVRHPDAMKSVFNADVSIKELAAIIOKDETSVTEKERIKELME 299
QY 325 SVEPPLVIAIKNIKRGLOALTQKSEVPLEPDPVOTOLDRQCEIRPGCVGVPAGGYDATA 384
DB 300 MNQGLILSGVSHSIEAV-----LITTVKHKLVSK-----LTGAGGGGCVL 341
QY 385 VLVLENQVGNFKOKTLE 401

```


PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157785.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161982.
PR	29-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query	March	5.5%	Score 123;	DB 21,	Length 380;
Best	Local Similarity	20.1%;	Pred. No. 0.02;		
Matches	88;	Conservative 74;	Mismatches 145;	Indels 130;	Gaps 17;
Oy	7	APGKAFLAGGLVLEPIYDAYVTALSSBMHVNIT---	EKGTSLKSRKIKISSPOFANQEW	63	
		: : : : : : : :			
Db	9	APGKIILGEEHNV---VHGSTRVAAHAIIDLYVTYVIRFLPLPSAENNDRIJTL---	QIKDITSL	62	
Oy	64	EYHHSNTEKREVOQRINPLEAETITVLAVIQTEAFDEIILISDPGYHSGE-----		118	
		: : : : : : : : : :			
Db	63	EFSMS-----LARIKAIPIYDSITLGRSPFACSBETLAKSI	98		
Oy	119	-----DTEYKTSNGKEFTLY-HSRAT-----TEVE-KTGLSSAGLVSV	157		

```

Dd      99  AULVERQONLPKREKMMLSGSIPTFLWLYTRIIIGFNPAIVVINSLEPYSGGLASSNALCVL 158
Qy      158  ATSLSHFT-----PNVISTKDIILHNVAOIAHCYAOKI-----GSGFDVATAIY 204
Dd      159  TALLASSISEKTRGNGMSGSLDETLYLELNKKA-----FEGEKIIHGKPSGIDWTSAYG 213
Qy      205  LIYRRFQFALINDPVOYLESDEPEKFPPELKKLIESNMBEGHECTLPYGIKILMGVKG 264
Dd      214  NML-----FEGSEITRLQSN-----MP-LRMLITNTRV 241
Qy      265  GSETPKLVSRYLQWKEKEPEBESSVYVYDOLANSANIQMFKELREMEKEKDSDEPYIKELDH 324
Dd      242  GRRTKALVSGVQRAVRHRPDAMKSIFNNAVDSIKEMALIIOSKDSVTSEKEKRIKELME 301
Qy      325  SVEPLTVAIKNIRKGLQALTOKSEVPIEPDVQOTOLBDCOEIPEGCVGVPVAGGYDAI 384
Dd      302  MNQGLLSMGVSHSSIEAV-----ILTYKHKLVS*-LTGAGGGGCVL 343
Qy      385  VLVLENQVGNFQKOTLE 401
Dd      344  TLL--PTGVVDKAVE 357

```

RESULT 28	
AAG50089	
ID	AAG50089 standard; Protein; 398 AA.
XX	
AC	AAG50089;
XX	
DT	18-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 63437.
XX	
KW	Protein identification; signal transduction pathway; metabolic pathway;
KM	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	termination sequence.
XX	
XX	Arabidopsis thaliana.
OS	
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134256.
PR	11-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145218.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 28-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149375.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151428.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157865.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 13-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160747.
 PR 21-OCT-1999; 99US-0160766.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161408.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.4%; Score 123; DB 21; Length 398;
Best Local Similarity 20.1%; Pred. No. 0.022;
Matches 88; Conservative 74; Mismatches 145; Indels 130; Gaps 17;

QY 7 APGKATLACGYLVLEPTDYAVYALSRMAVIT---PKQTSLKESRIKISSQFANGEW 63
DB 27 APKILLAGHAAV---VHGSTAAVAADLYTYVTLAFPLPSAANNDRLLT---QLKDLSL 80
QY 64 EYHISNTEKPREVQASINPFLATIFVLAAYQPTFAPELLEIIYSDPGHSGF----- 118
DB 81 EFSWS-----LARIKALPYDSSTLCSTSPASCSEETLKST 116
QY 119 -----DTEKTSNGEKTFLY-HSRAT-----TEVE-KTGLSGAGLVSVV 157
DB 117 AVLVESONLPKEMWMLSSGISTLMLYTRITGPNPAVIVNSLPIYSGSGSSALCLVAL 176
QY 158 AITSLSHFI-----PNVISTNKDILHVAQIAHCVAOKKI---GSGFDVATAIYG 204
DB 177 TALLASSISEKTRGNGWSLDETNLELLNKMA-----FEGEKIHKPGIGINDTVSAVG 221
QY 205 LIYRRFRQPLINDVQVLESDPEKPTTELKLIESWKEKHRCITLPYIKLMDVYG 264
DB 232 NMT-----KFCSGITRLQSN-----MP-LRMLITNTRV 259
QY 265 GSETPKLVSRYLQWKKEKPESSVVYDOLNSANILQFEKELREKRYSDPEYIKELDH 324
DB 260 GRNTKALVSGVSGRAVRHPAMKSVFNVAVDSIKELAAIIQSDDEFSVTEKERIKELWE 319
QY 325 SVEPLVAIKNIKGIALTQKSEVPREPVOYQOLDRCOEIPGCVGVVPGAGYDAA 384
DB 320 MNOGLLIMSGVSHSIEAV-----ILTVKHKLVSK-----LTGAGGGGCVL 361
QY 385 VLVLENQVGNFKOKTLE 401
DB 362 TLL---PTGTVDKVE 375

RESULT 29

AAGS0074 standard; Protein; 447 AA.

AC AAGS0074;
XX
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63416.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX BP1033405-A2.
PN
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128334.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0136229.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145092.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147392.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149802.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.5%; Score 123; DB 21; Length 447;
Best Local Similarity 20.1%; Pred. No. 0.026;
Matches 88; Conservative 74; Mismatches 145; Indels 130; Gaps 17;

QY 7 AGRKAFLAGGVLYLEPIDAYVTALSRMHAVIT---PKGSIKSRKIKISSPORGANGW 63
DB 76 AFGKIIILAGEHAV--VHGSTRVAAADLYTYVTIRFPLPSKENDRLTL--QKDIISL 129
QY 64 EYHISNTEKPREVOGRINPLEATIPIYAVIOPTFAFDLEIIYSPGYHSOR----- 118
DB 130 EFSWS-----LARIKAIYDSTLCRSTPASCSEETLSI 165
QY 119 -----DETETKSNGEKTFYV-HSRAT-----TEVE-KTGLSSAGIYSVV 157
DB 166 AVLVEEQNLPEKCKWMLSSGISITFELVYTRIGFNPATVIVINSELPGGLSSALCVAL 225
QY 158 ATSLISHFI-----PNVISTKDIHANVAQIAHCYAQKTI---GSGFDVATAIYG 204
DB 226 TAAIIASSISEKTRGNWSSIDETNLIELINKWA-----FEGEKTIHGKPSGIDNTVSAYG 280
QY 205 LIIYRRFPQPALINDFVYLESDEPEKPETELKKLIESNMEEKHERCTLPYGIKLIMGDVKG 264
DB 281 NMI-----KFCGSEITRLOS-----MP--LRLMIIITNRV 308
QY 265 GSETPKLVSRVYLQWKEKPESSVYVYDQNSANIQFMKELBMRKRYSDPEYIKELDH 324
DB 309 GRNTKALVSGVQRAVRHPRDAMKSVFNNAVDSIKELAIIQSKDSTVTEKEERIKELMB 368
QY 325 SVEPIITVAIKIRIKKQLQLTQKSEVPIPPDQTOQLDCCQETPGCVGVVPRGAGYDAIA 384
DB 369 MNQGLLSMGVSHSIEAV-----ILTTVKHKLIVSR-----LTGAGGGGCVTL 410
QY 385 VLVLENGVGNFKOKTLE 401
DB 411 TLL---PIGTVDKAYE 424
RESULT 30
AAG50073
ID AAG50073 standard; Protein, 451 AA.
XX AAG50073;
AC
XX
DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 63415.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
PN
XX 06-SEP-2000.
PD
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132483.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134378.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136392.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138867.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0144297.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.


```
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
Query Match 5.5%; Score 123; DB 21; Length 451;
Best Local Similarity 20.1%; Pred. No. 0.027;
Matches 88; Conservative 74; Mismatches 145; Indels 130; Gaps 17;
```

```
QY 7 ARGKAFACGYLYLEITYAYTALSSRMHAYIT--PKGTSLKESRIKISSPPQANGEW 63
DB 80 AGCKIILAGEHAYV--VHGSTVAALIDYTYVTTLRFPPLPSAENDRLTL--QLKDISL 133
QY 64 EKHISNTEKPREVSRINPFLFATIFVIAYIQPEAFDLIIYSDPGHSGF----- 118
DB 134 EFSWS-----LARIKEAIPYDSSTLCRSTPASCSSEETLSKI 169
QY 119 -----DTEKTSNGEKTFLY-HSRPI-----TEVE-KTGLGSSAGIVSV 157
DB 170 AVLVREQONPKKEMMLSSGISITFIMLYTIIIGENPATVYINSELPYGSGLSSALCVYL 229
QY 158 ASSLASHFT-----ENVISTNKDIIAHNAQIAHCYAOKKI---GSGFDVATAIYG 204
DB 230 TAAALASSISEKTRGNGWSLDETINLELNKWA-----FEGEKIHKPSGIDNTVSAVG 284
QY 205 LVIYRFFQALINDVQVLESDEKPEPTELKLISSNWEKKEKRCRLPYGICILMGDVYKG 264
DB 285 NMI-----KFCSEITRLQSN-----MP-LMMLITNTRV 312
```

```
QY 265 GSEPTPLVSRVLQWKKEKEPEESSVYDOLNSANLQFMKEIRREKRYDSDPETIKELDH 324
DB 313 GANTKALVSGVQRAVRHEDAKSVFNVAVDLSISKELAIATQKDETSVTEKEKRIEELNE 372
QY 325 SVEPLTVAIKIRKGLQALTQKSEVPIEDPVOTQLDRQCEIIPGCVGVPPAGGYDAIA 384
DB 373 MNQGLLSWGVSHSSIEAV-----LITTVGHKLVSX-----LTGAGGGGCVL 414
QY 385 VAVLENQNGNFKOKTLE 401
DB 415 TLE---PTGTVVDKXVE 428
RESULT 31
AAG50072
ID AAG50072 standard; Protein; 456 AA.
AC AAG50072;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 63414.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridization assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0128845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
```

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150567.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.5%; Score 123; DB 21; Length 456;
Best Local Similarity 20.1%; Pred. No. 0.027;

KM Bacterial meningitis; pneumonia; sepsis; otitis media;
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;
 XX auditory; respiratory; gene therapy; vaccine.
 OS Streptococcus pneumoniae type 4 strain.
 PN WO200277021-A2.
 PD 03-OCT-2002.
 PF 27-MAR-2002; 2002WO-1B02163.
 XX 27-MAR-2001; 2001GB-0007658.
 PR (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX Masignani V, Tettelin H, Frazer C;
 PI Masignani V, Tettelin H, Frazer C;
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 XX due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
 PT or ear infection -
 PS Claim 1; SEQ ID No 602; 56bp; English.
 XX
 PS The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC AB566454. Also included are an antibody which binds one of the
 CC proteins, treating a patient by administering the protein. DNA or
 CC antibody (in a composition), a kit comprising first and second primers,
 CC which are the nucleic acid cited above or fragments between nucleotides
 CC 8-100 of a sequence not defined in the specification, for amplifying a
 CC target sequence contained within a Streptococcus nucleic acid sequence,
 CC where the first primer is substantially complementary to the target
 CC sequence and the second primer is substantially complementary to the
 CC complement of the target sequence, and where the parts of the primers
 CC having substantial complementarity define the termini of the target
 CC sequence to be amplified, assay comprising contacting a test compound
 CC with the protein, and determining whether the test compound binds to the
 CC protein and a Streptococcus pneumoniae bacterium, where one or more
 CC genes encoding the proteins has been rendered inactive. The proteins,
 CC nucleic acid molecules, antibody and compositions are useful as
 CC medicaments for treating or preventing a disease or infection due to
 CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
 CC sepsis, otitis media or ear infection. They are also useful in developing
 CC vaccines, diagnostics and antibiotics. The methods are useful for
 CC identifying immunodominant proteins. The present sequence is one of
 CC the 2469 proteins expressed by the identified coding regions from the
 CC genomic sequence.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPD at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 335 AA;
 Query Match 5.2%; Score 117; DB 24; Length 335;
 Best Local Similarity 22.4%; Pred. No. 0.06;
 Matches 95; Conservative 53; Mismatches 146; Indels 130; Gaps 20
 QY 9 GRAFLAGGLVLEPIYDAYVATLSSRMAVITPKGTSLES-RIKISSOPANGEWEXHI 67
 Db 8 GGLYMAAGEVALLIEPGLALIKIDIPYRAEI-----AFSDSRIYSDMWDF----- 54
 QY 68 SSNTKEPREVOGRINP---FLAATFTIVLAIV---QTEAFDLEITIIYSDPGVHSQEDT 120

[illegible]

[illegible]

XX	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
FR	23-MAR-2000; 2000US-191637P.
FR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PMD, Myers EW;
XX	
DR	WPI; 2001-656860/75.
DR	N-PSDB; ABL16039.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	
XX	Disclosure, SEQ ID NO 42600; 21pp + Sequence listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC	sequences (ABB57737-ABB72072) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_ptc_sequences .
XX	
XX	Sequence 507 AA;

Query Match	5.0%	Score 110.5	DB 22	Length 507
Best Local Similarity	20.5%	Pred. No. 0.45		
Matches	92	Conservative 54	Mismatches 146	Indels 157
				Gaps 19
Qy	2	SKAAPSAGKAFLAGVLYLEPTIDAYVTALSSRMHNAVTPKGTSLKESRKISSPQFANG	61	
Db	10	SKKKSAPGIGFPIAG-----STNGSGIKIGL-VSSGTQ--NV	43	
Qy	62	EWETHISSNTEKP--REVQSHINPFLAATPIYLAVIQPTPEAFDLIIIIYSDPGHSGE	118	
Db	44	TSSQDISKRTKRPMEKRRRARIINQSLAILALIL-----E	79	
Qy	119	DETETKSSNGEKTPLHSAIIEVEKIGLSSAGLVSVATSLSHFIPNVISNK----	174	
Db	80	STYQNAKNGEQQAKH-----TKLEK-----ADLELTFRHFRHRNLDPTPNKTRAG	128	
Qy	175	--DILHNVAQIHACVQAOKKIGSGFVATAIYGLIYRRFPQALINDVFOVLSDPEKFPPT	232	
Db	129	YTDCAAEVARYLATPEPPMGT-----MPL-----AEPGSKAR	162	
Qy	233	ELKKLIESNWEKHERCTLPGYIKILMDVGKSGSETPLYS--RYLQMKKEKPESSVVY	290	
Db	163	LRLHNL-----DQCIAEIDVEICPSHTAFAFSPSSCPDINRGKKSQPREHSLDY	213	
Qy	291	DQLSANLQPMKELREMKRSDSDETIYIKEL-----DHSVEPLTVA--IKNIRKGIQAL	343	
Db	214	-----SSQDSNVDSKGLKRVAAEQRTLLPTPAPODENNRGLQAO	255	
Qy	344	TQSEVPIEPDVOTOLDRQCEIIPCCGVGVVPGAGGYDAIIVLLENOVGNFKQTKLENP	403	
Db	256	AQ--TPIPIQVQSGTQ-----GQTSPI--VAVAVAPSELSYEDRNKKCANVLEQYKQ-----	303	
Qy	404	DYFNHNYWDLLEQTEGVLEEKREDYIGL	432	
Db	304	-----QLKAHVQQQPFESANGV	319	

ID ABG06505 standard; Protein; 2633 AA.
AC ABG06505;
XX
XX
XX 13-FEB-2002 (first entry)
DE Novel human diagnostic protein #6496.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001MO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX N-PSDB; AAS70692.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
XX Claim 20; SEQ ID NO 36864; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2633 AA;
SQ

Query Match 4.9%; Score 108.5; DB 22; Length 2633;
Best Local Similarity 20.7%; Pred. No. 8.8;
Matches 84; Conservative 67; Mismatches 161; Indels 93; Gaps 16;

QY 47 KSRRIKISSPQANQWHEHISNTEKPREVQSRINPFLAIFLYAVIOETAPDFEI 106
DB 852 KRAQKFDSSIGALKTELSTYKQLOKTRIVERLNE-NEOUKEQLENDSPLOQVERBK 910
QY 107 ILYSDPGYHSDDETKTSSNGEKTFLYHSRAITVEKTKLGSSA----- 151
DB 911 TLTEKLOQTLEBVTITQ---EKODIKQLOESLQIEROQLKSDHDHYTNMNIIDQEOQLR 967
QY 152 -GLVSV---VATSLSHFTPNVISTNKDILHNVAAQIAHCVAOKKIG-----SGFD 197

DB 968 NALBSLKQHOETINTLKSISSEVRNLMMENTGETKQEFQOKWGDIDKQDLBAKNTQ 1027
QY 198 VATA-----IYGLIVRRFQPALINDPVQVLES---DPEKPTLEKLISEN 241
DB 1028 TLTAIVKONELLIEQQRKIFSLIQEK-----NELQOMLESYIAKEQDLTKLENIENT 1080
QY 242 WEKHERCTLPYGIKTLMDVDVKGSETPRLVSRVQWKK---EKPESSVVVVDOLNSANL 298
DB 1081 IENQSE-----IKLL-----GDELKKQDEIVAOEKHAIKGEGLSRTCDRLAEVBE 1127
QY 299 QPMKELEMRERKRYDS--DEPTYIKELDHSVEPLTVAINRKGLQALTKQSEVPTEPPVQ 356
DB 1128 KUKERKQQLQEKQQLLVNQESSEMOKKIN-----EINLKNEL-----KNNEITLE-HME 1178
QY 357 TOLLDRQGEIPGCVGVVPAGAGYALAVLVLENQVGNFKQKLE 401
DB 1179 TERLEAQKL-----NENYEVKSIPTERKVLKELQKSPFE 1213

RESULT 39
AAM39097
ID AAM39097 standard; Protein; 2663 AA.
XX
XX AAM39097;
XX
XX 22-OCT-2001 (first entry)
XX
XX
XX Human polypeptide SEQ ID NO 2242.
XX
XX
XX Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
KM peripheral nervous system; neuropathy; central nervous system; CNS;
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KM chemokine; chromolytic; drug screening; arthritis; inflammation;
KM leukaemia.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao Q, Zhou P, Goodrich R, Dmanac RT;
XX
XX WPI; 2001-442253/47.
XX N-PSDB; AAI58253.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 4; SEQ ID NO 2242; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with noctropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous

CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemoclastic/chemokine activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

CC Sequence 2663 AA;

Query Match 4.9%; Score 108.5; DB 22; Length 2663;

Best Local Similarity 20.7%; Pred. No. 9;

Matches 84; Conservative 67; Mismatches 161; Indels 93; Gaps 16;

CC 47 KESRIKISSPPFANGEMWEYHISSTNTEKREVRQSRINPFLKATTFIVLAVIQTEAFDLEI 106

CC 852 KEAQKFDSSLGALKTSLSTYQLOEKTRVEQERLNE-MEQLKEQLNRSPLQTVREK 910

CC 107 IIVSDPGYHSODETETKTSNGEKTFLYHSRAITEVEKTGLGSSA----- 151

CC 911 TLITTEKLOQTLEEVKTLTQ---EKDILKQLOESLQIERDQLKSDIHDTVMNIDTQEQRL 967

CC 152 -GLVSV---VATSLSHFIPNVISTNKDILHNVAQIAHCYAAKKIG-----SGPD 197

CC 968 NALLESKQHQETINTLTKSKISEVSRNLHMEENTGTEKDFEQQKMGIDKKQDLKAKNTQ 1027

CC 198 VATA-----IYGLIYRRFPQALINDVEQVLES---DPEKPTLEKLIEN 241

CC 1028 TLTAADVKNENIIEQQRKIFSLIOEK-----NELQOMLESVIAAEKQDKTDKENIEMT 1080

CC 242 WEEKHRCITLPGIKILMGDVKGSSETPKLVSRVLQWK---EKPESSVVVYDOLNSANL 298

CC 1081 IENQEE-----LRLI-----GDELKQOEIVAOEKNAIKKEGELSRTCRLAVERE 1127

CC 299 QFMKEIREMERKXDS--DPEYIKELDHSVEPLTVAIKNIRKGLQALTQKSEVPPIEPDVQ 356

CC 1128 KLEKESQQLQOEKQOQLNVOEEMSEMOCKIN---ELENKKNEL---KNKELTLE-HME 1178

CC 357 TOLLDRCOEIPGCVGVPAGGYDAIYAVLENQVGNFKQKTL 401

CC 1179 TERLELAQKL-----NENYEEVKSITTKERKVLKELQKSF 1213

CC RESULT 40

CC ID AAM40883

CC ID AAM40883 standard; Protein; 2688 AA.

CC AC AAM40883;

CC 22-OCT-2001 (first entry)

CC Human polypeptide SEQ ID NO 5814.

CC XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

CC XX peripheral nervous system; neuropathy; central nervous system; CNS;

CC XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

CC XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemoclastic;

CC XX chemokine; thrombolytic; drug screening; arthritis; inflammation;

CC XX leukemia.

CC XX Homo sapiens.

CC XX OS

CC XX PN WO200153312-A1.

CC XX PD 26-JUL-2001.

CC XX PF 26-DEC-2000; 2000WO-US34263.

CC XX PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.

XX (HYSR-) HYSEQ INC.

XX PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR MPI: 2001-442253/47.

XX DR N-PSDB; AA160039.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders

XX PT such as central nervous system injuries -

XX PS Example 2; SEQ ID NO 5814; 10078pp; English.

CC The invention relates to human nucleic acids (AA157796-AA161369) and
 CC the encoded polypeptides (AAM3642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemoclastic/chemokine activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

CC Sequence 2688 AA;

CC Query Match 4.9%; Score 108.5; DB 22; Length 2688;

CC Best Local Similarity 20.7%; Pred. No. 9.1;

CC Matches 84; Conservative 67; Mismatches 161; Indels 93; Gaps 16;

CC 47 KESRIKISSPPFANGEMWEYHISSTNTEKREVRQSRINPFLKATTFIVLAVIQTEAFDLEI 106

CC 876 KEAQKFDSSLGALKTSLSTYQLOEKTRVEQERLNE-MEQLKEQLNRSPLQTVREK 934

CC 107 IIVSDPGYHSODETETKTSNGEKTFLYHSRAITEVEKTGLGSSA----- 151

CC 935 TLITTEKLOQTLEEVKTLTQ---EKDILKQLOESLQIERDQLKSDIHDTVMNIDTQEQRL 991

CC 152 -GLVSV---VATSLSHFIPNVISTNKDILHNVAQIAHCYAAKKIG-----SGPD 197

CC 992 NALLESKQHQETINTLTKSKISEVSRNLHMEENTGTEKDFEQQKMGIDKKQDLKAKNTQ 1051

CC 198 VATA-----IYGLIYRRFPQALINDVEQVLES---DPEKPTLEKLIEN 241

CC 1052 TLTAADVKNENIIEQQRKIFSLIOEK-----NELQOMLESVIAAEKQDKTDKENIEMT 1104

CC 242 WEEKHRCITLPGIKILMGDVKGSSETPKLVSRVLQWK---EKPESSVVVYDOLNSANL 298

CC 1105 IENQEE-----LRLI-----GDELKQOEIVAOEKNAIKKEGELSRTCRLAVERE 1151

CC 299 QFMKEIREMERKXDS--DPEYIKELDHSVEPLTVAIKNIRKGLQALTQKSEVPPIEPDVQ 356

CC 1152 KLEKESQQLQOEKQOQLNVOEEMSEMOCKIN---ELENKKNEL---KNKELTLE-HME 1202

CC 357 TOLLDRCOEIPGCVGVPAGGYDAIYAVLENQVGNFKQKTL 401

CC 1203 TERLELAQKL-----NENYEEVKSITTKERKVLKELQKSF 1217

Mon Aug 18 10:26:23 2003

us-10-069-062-7.rag

Page 35

Search completed: August 8, 2003, 09:37:24
Job time : 90 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: August 8, 2003, 09:40:37 ; Search time 83 Seconds
(without alignments)
826.143 Million cell updates/sec

Title: US-10-069-062-7

Perfect score: 432

Sequence: 1 MSKAFSAPGKAFLAGYLV.....DLEQTEGVLEKEPDYIGL 432

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : A.Geneseq 19Jun03.*

```
1: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	432	100.0	432	22	AAV72679
2	179	41.4	432	22	AAU15094
3	179	41.4	432	23	ABP73186
4	13	3.0	451	24	ABJ25748
5	13	3.0	484	24	ABJ26348
6	11	2.5	451	23	ABP53757
7	8	1.9	234	21	AA651913
8	8	1.9	307	21	AA651912
9	8	1.9	357	22	AAU61184

10	8	1.9	376	21	AA651911	Arabidopsis thaliana
11	8	1.9	503	24	AAO27025	Acetyl-CoA acetyl
12	8	1.9	505	23	ABP77505	Arabidopsis thaliana
13	8	1.9	1378	24	ABP71102	Murine homologue of
14	8	1.9	1379	24	ABP71087	Human homologue of
15	7	1.6	56	22	AAU63540	Protonibacterium
16	7	1.6	59	22	AAU60558	Protonibacterium
17	7	1.6	77	23	ABP34327	Human ORF3300 prot
18	7	1.6	86	20	AAU08017	Mouse partial TAP
19	7	1.6	90	22	AAU87913	Human immune/haema
20	7	1.6	108	22	AAU36238	Pseudomonas aerugi
21	7	1.6	108	24	ABJ18790	Pseudomonas aerugi
22	7	1.6	109	21	AAU40712	Zea mays protein f
23	7	1.6	115	22	AAU34293	Peptide #8330 enco
24	7	1.6	124	21	AAU36977	Arabidopsis thaliana
25	7	1.6	132	20	AAU36977	Amino acid sequenc
26	7	1.6	137	22	ABP59022	Drosophila melanog
27	7	1.6	146	22	AAU81086	Human haematologic
28	7	1.6	146	22	AAU81536	Human haematologic
29	7	1.6	149	23	ABU61012	Lung specific prot
30	7	1.6	157	23	ABP64574	Human ORF944. Hom
31	7	1.6	176	21	AAU84688	Human acid sequenc
32	7	1.6	176	22	AAU11054	Ashbya gossypii ph
33	7	1.6	178	21	AAU40711	Zea mays protein f
34	7	1.6	179	23	ABP44892	Human protein SBQ
35	7	1.6	183	21	AAU34181	Zea mays protein f
36	7	1.6	184	22	AAU31841	Human lung tumour-
37	7	1.6	186	23	ABU05936	M. tuberculosis an
38	7	1.6	189	23	ABG68071	Human interferon a
39	7	1.6	189	23	ABG68076	Human interferon a
40	7	1.6	193	23	ABP42703	Human ovarian anti
41	7	1.6	204	22	AAU13875	Human polypeptide
42	7	1.6	204	23	AAU69528	Human G protein-co
43	7	1.6	220	15	AAU60770	Maize OM homolog f
44	7	1.6	220	21	AAU34110	Zea mays protein f
45	7	1.6	225	23	ABP73992	Candida albicans e

ALIGNMENTS

RESULT 1	AAV72679	standard; Protein; 432 AA.
ID	AAV72679	
AC	AAV72679;	
DT	31-MAY-2001 (first entry)	
DE	Candida albicans phosphomevalonate kinase (PMK, ERG8) protein.	
DE	Phosphomevalonate kinase; PMK; ERG8; anti-fungal agent; diagnosis; infection.	
KW	Candida albicans.	
OS	Candida albicans.	
XX	Key	Location/Qualifiers
XX	Misc-difference 244	/note="Encoded by TTC of the sequences shown in
FT	AAU02791 and AAU02792"	
FT	AAU02791 and AAU02792"	
PN	MO200114533-A2.	
PD	01-MAR-2001.	
XX	15-AUG-2000; 2000WO-GB03100.	
PF	21-AUG-1999; 99GB-0019766.	
PR	(ASTR) ASTRAZENCA AB.	
XX	(ASTR) ASTRAZENCA UK LTD.	
PA	(ASTR) ASTRAZENCA UK LTD.	
XX	Rosamond JDC, Schnell NF;	
XX	PI	

XX WPI; 2001-218441/22.
DR N-PSDB; AAD02791, AAD02792.
XX
PT New polypeptides and polynucleotides (ERG8) from *Candida albicans*,
XX useful in assays for identifying inhibitors of phosphomevalonate kinase
XX activity and as reagents for diagnosing *C. albicans* infection -
XX
PS Claim 1; Page 27-28; 29pp; English.
CC The present sequence is phosphomevalonate kinase (PMK; ERG8) protein
CC from *Candida albicans*. The ERG8 protein is useful in an assay for
CC identifying compounds that inhibit phosphomevalonate kinase (PMK)
CC activity. These inhibitors are useful as anti-fungal agents. The ERG8
CC DNA and protein are also useful as reagents for diagnosing *C. albicans*
CC infection.
XX
SQ Sequence 432 AA;
Query Match 100.0%; Score 432; DB 22; Length 432;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSKAFAPGKAFLAGVLYLEPIYDAVYVALSSRMNAVITPKGTSLKESRIKISDPQAN 60
DB 1 MSKAFAPGKAFLAGVLYLEPIYDAVYVALSSRMNAVITPKGTSLKESRIKISDPQAN 60
QY 61 GEMEHYISSNTEKPREVQSRINPFLNATIFIVLAIQPTFAPPLEIIYSDPGYHQEDT 120
DB 61 GEMEHYISSNTEKPREVQSRINPFLNATIFIVLAIQPTFAPPLEIIYSDPGYHQEDT 120
QY 121 ERTSSNGEKTFPLYSRATIEVEKGTGSSAGLVSVATSLSHFIPNYISTKDLHNHY 180
DB 121 ERTSSNGEKTFPLYSRATIEVEKGTGSSAGLVSVATSLSHFIPNYISTKDLHNHY 180
QY 181 AQIAHCYAOIKTIGSGFDVATAIYGLIVRRFPQALINDVFQVLESDEPEKTELKLI 240
DB 181 AQIAHCYAOIKTIGSGFDVATAIYGLIVRRFPQALINDVFQVLESDEPEKTELKLI 240
QY 241 NMEEKHERCTLPVIGILMGDVKGSETPKLVSRVQMKKEKEESSVYVDQNSANLQF 300
DB 241 NMEEKHERCTLPVIGILMGDVKGSETPKLVSRVQMKKEKEESSVYVDQNSANLQF 300
QY 301 MEELRMRERKYSDDPEYIKELDHSEVPLTVAIKNIRKGLQALTQKSEVPIDPVQTL 360
DB 301 MEELRMRERKYSDDPEYIKELDHSEVPLTVAIKNIRKGLQALTQKSEVPIDPVQTL 360
QY 361 DRQCEIPGCVGVVPGAGGYDAIAVLLENQVGNFKQKTLNPDYFHNHYVVDLEQTEG 420
DB 361 DRQCEIPGCVGVVPGAGGYDAIAVLLENQVGNFKQKTLNPDYFHNHYVVDLEQTEG 420
QY 421 VLEEKPEDYIGL 432
DB 421 VLEEKPEDYIGL 432

RESULT 2
AAU15094
ID AAU15094 standard; Protein; 432 AA.
XX
AC AAU15094;
XX
DT 04-DEC-2001 (first entry)
XX
DE Protein encoded by *C. albicans* essential gene CayMR220W (ERG8).
XX
KW Gene identification; essential gene; GRACE; pathogenic fungus;
KW gene replacement and conditional expression; fungal infection.
XX
OS *Candida albicans*.
XX
PN WO200160975-A2.
XX

PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05551.
XX
XX 18-FEB-2000; 2000US-0183534.
XX
PA (ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H;
XX
XX WPI; 2001-489080/53.
DR N-PSDB; AAS23422.
XX
PT Identifying genes essential to fungal metabolisms and identifying
XX potential therapeutic agents that target these genes -
XX
PS Claim 43; Page 226-227; 324pp; English.
XX
CC The present invention relates to novel methods for constructing fungal
CC strains useful for identification and validation of gene products as
CC targets for therapeutic agents, for creating a collection of identified
CC essential genes, and screening assays for the discovery of new drugs.
CC The invention provides the GRACE (gene replacement and conditional
CC expression) method for the construction of mutant organisms referred to
CC as GRACE strains of the organism. The invention can be applied to any
CC organism, particularly a pathogenic fungus e.g. *Candida albicans*,
CC *Aspergillus fumigatus* and *Cryptococcus neoformans*. The methods are
CC useful to identify agents that may be used in the treatment of fungal
CC infections. AAU15053-AAU15113 represent proteins encoded by *C. albicans*
CC essential genes.
XX
SQ Sequence 432 AA;
Query Match 41.4%; Score 179; DB 22; Length 432;
Best Local Similarity 100.0%; Pred. No. 8.1e-166;
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 254 GIKLMGDKVKGSETPKLVSRVQMKKEKEESSVYVDQNSANLQFMEKELREMEKXDS 313
DB 254 GIKLMGDKVKGSETPKLVSRVQMKKEKEESSVYVDQNSANLQFMEKELREMEKXDS 313
QY 314 DEYTIKELDHSEVPLTVAIKNIRKGLQALTQKSEVPIDPVQTLDRQCEIPGCVGV 373
DB 314 DEYTIKELDHSEVPLTVAIKNIRKGLQALTQKSEVPIDPVQTLDRQCEIPGCVGV 373
QY 374 VPGAGGYDAIAVLLENQVGNFKQKTLNPDYFHNHYVVDLEQTEGVLEKPEDYIGL 432
DB 374 VPGAGGYDAIAVLLENQVGNFKQKTLNPDYFHNHYVVDLEQTEGVLEKPEDYIGL 432

RESULT 3
ABP73186
ID ABP73186 standard; Protein; 432 AA.
XX
AC ABP73186;
XX
DT 30-JAN-2003 (first entry)
XX
DE *Candida albicans* essential protein SEQ ID NO 7023.
XX
KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; *Candida albicans*; fungicide; antifungal.
XX
OS *Candida albicans*.
XX
PN WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US49486.
XX
PR 29-DEC-2000; 2000US-259128P.
XX

PR 20-FEB-2001; 2001US-0792024.
PR 22-AUG-2001; 2001US-314050P.
XX
XX
XX (ELIT-) ELITRA PHARM INC.
PI Roemer T, Jiang B, Boone C, Buseey H, Ohlsen KJ;
XX WPI; 2002-566694/60.
DR N-PSDB; AB231736.
XX
XX
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele
PT of a gene and placing other allele of the gene under conditional
XX expression -
XX
PS Claim 44; SEQ ID NO 7023; 167pp + Sequence Listing; English.
XX
XX The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans protein used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX
SQ Sequence 432 AA;
Query Match 41.4%; Score 179; DB 23; Length 432;
Best Local Similarity 100.0%; Pred. No. 8.1e-166;
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 254 GIKLMGVKGGSETPKLVSRVLOMKKEPSESSVYVDOLNSANTQFMKELREMEKTD 313
DB 254 GIKLMGVKGGSETPKLVSRVLOMKKEPSESSVYVDOLNSANTQFMKELREMEKTD 313
QY 314 DEETVYKEIDHSVEPLVTAIKIRKGLQALTKXSEVPIEPDVQTLDRCCQIPGCVGV 373
DB 314 DEETVYKEIDHSVEPLVTAIKIRKGLQALTKXSEVPIEPDVQTLDRCCQIPGCVGV 373
QY 374 VPGAGGYAIAVLVLENQVGNKQKTLEPDYFNHVVWVWDLSEQTEGYLEEPEDYIGL 432
DB 374 VPGAGGYAIAVLVLENQVGNKQKTLEPDYFNHVVWVWDLSEQTEGYLEEPEDYIGL 432
RESULT 4
ABJ25748
ID ABJ25748 standard; Protein; 451 AA.
XX
XX ABJ25748;
AC
XX
XX 16-APR-2003 (first entry)
DT
XX
XX Aspergillus fumigatus essential gene protein #406.
DE
XX
XX Fungicide; cytosstatic; essential gene; Aspergillus fumigatus; infection;
KW

KW cancer; contamination; biofilm; antibody; immune response.
XX
XX Aspergillus fumigatus.
OS
XX
XX WO200286090-A2.
PN
XX
XX 31-OCT-2002.
PD
XX
XX 23-APR-2002; 2002MO-US13142.
PF
XX
XX 23-APR-2001; 2001US-285697P.
PR 27-APR-2001; 2001US-287066P.
PR 05-JUN-2001; 2001US-295890P.
PR 09-JUL-2001; 2001US-303899P.
PR 31-AUG-2001; 2001US-316362P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
PI WPI; 2003-093124/08.
DR
XX
XX
XX New purified or isolated nucleic acids of essential genes of
PT Aspergillus fumigatus, useful for treating or preventing infections by
PT A. fumigatus, or for treating a non-infectious disease in a subject
PT e.g. cancer -
XX
PS Disclosure; Page -; 175pp; English.
XX
XX The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or inhibit formation of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case of virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention.
XX
SQ Sequence 451 AA;
Query Match 3.0%; Score 13; DB 24; Length 451;
Best Local Similarity 100.0%; Pred. No. 0.00089;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 371 GGVVPGAGGYDAI 383
DB 379 GGVVPGAGGYDAI 391
RESULT 5
ABJ26348
ID ABJ26348 standard; Protein; 484 AA.
XX
XX ABJ26348;
AC
XX
XX

DT	16-APR-2003	(first entry)
DE	Aspergillus fumigatus essential gene protein #1006.	
XX		
XX	Fungicide; cyrostatic; essential gene; Aspergillus fumigatus; infection;	
KW	cancer; contamination; biofilm; antibody; immune response.	
XX		
OS	Aspergillus fumigatus.	
XX		
PN	MO200286090-A2.	
XX		
PD	31-OCT-2002.	
XX		
PF	23-APR-2002; 2002WO-US13142.	
XX		
PR	23-APR-2001; 2001US-285697P.	
PR	27-APR-2001; 2001US-287066P.	
PR	05-JUN-2001; 2001US-295890P.	
PR	09-JUL-2001; 2001US-303899P.	
PR	31-AUG-2001; 2001US-316362P.	
XX		
PA	(ELIT-) ELITRA PHARM INC.	
XX		
PI	Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;	
XX	WPI, 2003-093124/08.	
XX		
PT	New purified or isolated nucleic acids of essential genes of	
PT	Aspergillus fumigatus, useful for treating or preventing infections by	
PT	A. fumigatus, or for treating a non-infectious disease in a subject	
PT	e.g. cancer -	
XX		
PS	Disclosure; Page -: 175pp; English.	
XX		
XX		
CC	The invention relates to novel purified or isolated nucleic acids of	
CC	essential genes of Aspergillus fumigatus. The isolated nucleic acids of	
CC	the invention are used to treat or prevent infections by a pathogenic	
CC	organism such as A. fumigatus, to treat a non-infectious disease in a	
CC	subject (e.g. cancer), to prevent or contain contamination of an object	
CC	by A. fumigatus, to prevent or inhibit formation on a surface of a	
CC	biofilm comprising A. fumigatus. The polynucleotides are useful for	
CC	expressing recombinant protein for characterisation, screening or	
CC	therapeutic use, as markers for host tissues in which the pathogenic	
CC	organisms invade, or reside, for comparing with the DNA sequence of A.	
CC	fumigatus to identify duplicated genes or paralogues having the same or	
CC	similar biochemical activity and/or function, for comparing with DNA	
CC	sequences of other related or distant pathogenic organisms to identify	
CC	potential orthologous essential or virulence genes, for selecting and	
CC	making oligomers for attachment to a nucleic acid array for examination	
CC	of expression patterns, for raising anti-protein antibodies, as an	
CC	antigen to raise anti-DNA antibodies or to elicit another immune	
CC	response, and for identifying polynucleotides encoding the other protein	
CC	with which binding occurs or to identify inhibitors of the binding	
CC	interaction. The polypeptides may be used to raise antibodies or to	
CC	elicit immune response, as a reagent in assays designed to quantitatively	
CC	determine levels of the protein in biological fluids, as a marker for	
CC	host tissues in which pathogenic organism invade or reside, and to	
CC	isolate correlative receptors or ligands in the case of virulence	
CC	factors. This sequence represents a protein of one of the essential genes	
CC	of Aspergillus fumigatus of the invention.	
XX		
SQ	Sequence 484 AA;	
XX		
Query Match	3.0%;	Score 13; DB 24; Length 484;
Best Local Similarity	100.0%;	Pred. No. 0.00095;
Matches 13; Conservative	0;	Mismatches 0; Indels 0; Gaps 0
QY	371	GGVPGAGGYDAI 383
DB	412	GGVPGAGGYDAI 424

ABP53757
ID ABP53757 standard; Protein; 451 AA.
XX
AC ABP53757;
XX
DT 03-JAN-2003 (first entry)
DE Saccharomyces cerevisiae mevalonate kinase SEQ ID NO:24.
KW Mevalonate kinase; enzyme; herbicide; agricultural; plant.
XX
OS Saccharomyces cerevisiae.
PN US2002119546-A1.
PM
PD 29-AUG-2002.
PF 20-JUL-2001; 2001US-0909745.
PR 05-NOV-1998; 98US-107241P.
PP 04-NOV-1999; 99US-0433242.
PA (FALC/) FALCO S C.
PI (FAMO/) FAMODU O O.
PT Falco SC, Famodu OO;
DR WPI, 2002-731357/79.
XX
PT New mevalonate kinase polypeptide useful for new herbicide discovery
and design, and for altering levels of mevalonate kinase in transformed
cells -
PS
Example 4; Fig 3; 39pp; English.

The present invention describes an isolated mevalonate kinase polypeptide
(I), having a sequence identity of 80 % based on the Clustal method of
alignment when compared to corn, rice, soybean and wheat mevalonate
kinase. (I) has herbicidal and agricutlural activities. A nucleic acid
(II) encoding (I) is useful for transforming a cell and producing a
transgenic plant. A chimeric gene (III) comprising (II) is useful for
altering the level of expression of a squalene biosynthetic enzyme in a
host cell, where the expression of the chimeric gene results in
production of altered levels of mevalonate kinase in the transformed host
cell. (III) is also useful for evaluating a compound for its ability to
inhibit the activity of a mevalonate kinase. (I) is useful as a target
to facilitate the design and/or identification of inhibitors of these
enzymes that may be useful as herbicides. Fragments of (II) are useful
to create transgenic plants in which the disclosed polypeptides are
present at higher or lower levels than normal or in cell types or
developmental stages in which they are not normally found. The nucleic
acid fragments are also useful as probes for genetically and physically
mapping the genes that they are a portion of and as markers for traits
linked to those genes. The present sequence represents a Saccharomyces
cerevisiae mevalonate kinase protein from the present invention.

Sequence 451 AA;

Query Match 2.5%; Score 11; DB 23; Length 451;
Best Local Similarity 100.0%; Pred. No. 0.08;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0

CY 144 KTGLGSSAGLV 154
|||||
DB 152 KTGIGSSAGLV 162

RESULT 7
AAGS1913
ID AAGS1913 standard; Protein; 234 AA.
NC
XX AAGS1913;
XX

DT 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 65934.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130649.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134219.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138547.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139453.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155349.
PR 24-SEP-1999; 99US-0155559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.9%; Score 8; DB 21; Length 234;
Best Local Similarity 100.0%; Pred. No.38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 232 TELKKLIE 239
 |||||
Db 3 TELKKLIE 10

RESULT 8
AAG51912
ID AAG51912 standard; Protein; 307 AA.
XX
AC AAG51912;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 65933.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0122180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144684.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145086.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145813.
 PR 27-JUL-1999; 99US-0145818.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.

PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155139.
 PR 24-SEP-1999; 99US-0155655.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161921.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.9%; Score 8; DB 21; Length 307;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 TELKKLIE 239
 Db 76 TELKKLIE 83

RESULT 9
 AAU61184
 ID AAU61184 standard; Protein; 357 AA.
 XX AAU61184;
 DT 27-FEB-2002 (first entry)
 DE Propionibacterium acnes immunogenic protein #22080.
 XX
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 OS Propionibacterium acnes.
 XX
 PN WO200181581-A2.
 XX
 PD 01-NOV-2001.

XX PF 20-APR-2001; 2001MO-US12865.
 XX PR 21-APR-2000; 2000US-199047P.
 XX PR 02-JUN-2000; 2000US-208841P.
 XX PR 07-JUL-2000; 2000US-216747P.
 XX PA (CORI-) CORIXA CORP.
 XX PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhactia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 DR MPI; 2001-616774/71.
 DR N-PSDB; AAS59615.
 XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 XX PS Example 1; SEQ ID No 22379; 10699P; English.
 XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 357 AA;
 Query Match 1.9%; Score 8; DB 22; Length 357;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 377 AGGYDAIA 384
 Db 60 AGGYDAIA 67
 RESULT 10
 ID AAG51911 standard; Protein; 376 AA.
 AC AAG51911;
 XX 18-OCT-2000 (first entry)
 DT Arabidopsis thaliana protein fragment SEQ ID NO: 65932.
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 65932.
 XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KM termination sequence.
 XX Arabidopsis thaliana.
 OS
 XX EP1033405-A2.
 PN
 XX 06-SEP-2000.
 PD
 XX

PF 25-FEB-2000; 2000EP-0301439.
 XX 25-FEB-1999; 99US-0121825.
 XX 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0128845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131445.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139464.
 PR 18-JUN-1999; 99US-0139465.
 PR 18-JUN-1999; 99US-0139466.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144362.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.

PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.9%; Score 8; DB 21; Length 376;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 232 TELKKLIE 239
|||||
Db 145 TELKKLIE 152

RESULT 11

AAO27025 standard; Protein; 503 AA.

XX AAO27025;

DT 15-MAY-2003 (first entry)

DE Acetyl-coA acetyltransferase enzyme protein, SEQ ID No 12.

XX Isopentenyl diphosphate: IPP; pathway enzyme; IPP biosynthesis;

KW acetyl-coA acetyltransferase enzyme; acetate.

XX Hevea brasiliensis.

OS WO2003010294-A2.

XX WO2003010294-A2.

PD 06-FEB-2003.

XX 23-JUL-2002; 2002WO-US24048.

XX 25-JUL-2001; 2001US-307673P.

PA (DUPO) DU PONT DE NEMOURS & CO E I.

XX Hallahan DL, Keiper-hrynko NM;

XX

DR MPI; 2003-239439/23.
 DR N-PSDB; AAL55372.
 XX
 PT Novel isolated nucleic acid molecule encoding isopentenyl diphosphate,
 PT IPP, pathway enzyme, useful for obtaining nucleic acid molecule
 PT encoding IPP pathway enzyme, and for regulating IPP biosynthesis in
 PT organism -
 XX
 PS Claim 4; Page 62-63; 66pp; English.
 CC This polynucleotide sequence represents an isolated nucleic acid molecule
 CC which encodes an isopentenyl diphosphate (IPP) pathway enzyme that has a
 CC 411, 464, 386, 503 or 415 residue amino acid sequence, given in the
 CC specification, hybridizes with nucleic acid molecule encoding the amino
 CC acid sequences, or is complementary to the sequences. The isolated
 CC nucleic acid is useful for regulating IPP biosynthesis in an organism,
 CC where the nucleic acid is overexpressed such that IPP biosynthesis is
 CC altered in the organism. The IPP pathway gene is over-expressed on a
 CC multicopy plasmid, and is operably linked to an inducible or regulated
 CC promoter. The IPP gene is optionally expressed in antisense orientation,
 CC or is disrupted by insertion of foreign DNA into the coding region. The
 CC isolated IPP nucleic acid or sequences showing identity are useful for
 CC obtaining nucleic acid molecules encoding IPP pathway enzymes, which
 CC involves probing a genomic library with the nucleic acid, identifying a
 CC DNA clone that hybridizes with the nucleic acid, and sequencing the
 CC genomic fragment that comprises the clone, where the sequenced genomic
 CC fragment encodes an IPP pathway enzyme. The isolated nucleic acid having
 CC a 1233, 1392, 1158, 1509 or 1245 nucleotide sequence, given in the
 CC specification, is useful for obtaining a nucleic acid molecule encoding
 CC an IPP pathway enzyme, which involves synthesizing at least one
 CC oligonucleotide primer corresponding to a portion of the sequence, and
 CC amplifying an insert present in a cloning vector using the
 CC oligonucleotide primer, where the amplified insert encodes a portion of
 CC an amino acid sequence encoding the enzyme. A transformed host cell is
 CC useful for producing a compound in the IPP pathway, which involves
 CC contacting a transformed host cell transformed with the isolated IPP
 CC nucleic acid under the control of suitable regulatory sequences, under
 CC suitable growth conditions with a carbon substrate, thus a compound in
 CC IPP pathway is produced. This sequence represents an acetyl-CoA
 CC acetyltransferase enzyme protein, a variant of one of the enzymes used to
 CC synthesize IPP from acetate.
 CC
 SQ Sequence 503 AA;
 XX
 Query Match 1.9%; Score 8; DB 24; Length 503;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 144 KTGLGSSA 151
 |||||
 DB 180 KTGLGSSA 187
 |||||
 RESULT 12
 ID ABB77505 standard; Protein; 505 AA.
 XX ABB77505;
 XX
 DT 29-JUL-2002 (first entry)
 XX
 DE Arabidopsis thaliana PMVK SEQ ID NO 2.
 XX
 KM Thale cress; PMVK; phosphomevalonate kinase; plant; herbicide;
 KM growth regulator; enzyme.
 XX
 OS Arabidopsis thaliana.
 OS
 PN DE10057755-A1.
 XX
 PD 23-MAY-2002.
 XX
 PF 22-NOV-2000; 2000DE-1057755.

XX
 PR 22-NOV-2000; 2000DE-1057755.
 XX
 PA (FARB) BAYER AG.
 XX
 PI Weisener R, Lechelt-Kunze C;
 XX
 DR MPI; 2002-445360/48.
 DR N-PSDB; ABL60244.
 XX
 PT New nucleic acid encoding plant phosphomevalonate kinase, useful for
 PT identifying modulators, potentially useful as herbicides and growth
 PT regulators -
 XX
 PS Claim 14; Page 13-14; 18pp; German.
 CC The invention relates to a nucleic acid (I, ABL60244) that encodes a
 CC plant phosphomevalonate kinase (PMVK, ABB77505), excluding the known
 CC fully defined partial sequences ABL60245-ABL60247. Plant PMVK, (I),
 CC constructs and host cells that contain (I) are used to identify agents
 CC that bind to and/or modulate activity of PMVK, potentially useful as
 CC herbicides and growth regulators. (I) is also used for recombinant
 CC production of PMVK.
 CC
 SQ Sequence 505 AA;
 XX
 Query Match 1.9%; Score 8; DB 23; Length 505;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 144 KTGLGSSA 151
 |||||
 DB 179 KTGLGSSA 186
 |||||
 RESULT 13
 ID ABB771102 standard; Protein; 1378 AA.
 XX ABB771102;
 AC
 XX
 DT 14-APR-2003 (first entry)
 XX
 DE Murine homologue of attractin/mahogany (HAM) polypeptide.
 XX
 KM HAM; homologue of attractin/mahogany; immunosuppressive; cytostatic;
 KM antiinflammatory; cardiant; osteopathic; gene therapy; mouse.
 XX
 OS Mus sp.
 OS
 PN WO200297120-A1.
 XX
 PD 05-DEC-2002.
 XX
 PF 23-MAY-2002; 2002WO-US16391.
 XX
 PR 25-MAY-2001; 2001US-293608P.
 PR 24-SEP-2001; 2001US-324626P.
 XX
 PA (IMMV) IMMUNEX CORP.
 PA
 PI Anderson DM;
 PI
 XX
 DR MPI; 2003-140486/13.
 DR N-PSDB; ABE58710.
 XX
 PT New Homologue of Attractin/Mahogany (HAM) polypeptide, useful for
 PT treating HAM-associated disorder consisting of inflammatory,
 PT autoimmune, cell proliferative or cardiovascular disorders -
 XX
 PS Claim 3; Fig 5A-D; 89pp; English.
 XX
 CC The invention relates to Homologue of Attractin/Mahogany (HAM)

CC polypeptides and encoding polynucleotides. The HAM polypeptides can be
 CC expressed by standard recombinant methodology. The HAM polypeptides are
 CC useful for treating HAM-associated disorder consisting of inflammatory,
 CC autoimmune, graft-versus-host, neurological, myelination, cell
 CC proliferative, cardiovascular, haematologic, liver, metabolic, weight or
 CC bone disorder. The present sequence represents a mouse HAM polypeptide.
 XX

Sequence 1378 AA;

Query Match 1.9%; Score 8; DB 24; Length 1378;

Best Local Similarity 100.0%; Pred. No. 1.8e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 EYHISNT 71
 |||||
 445 EYHISNT 452

Db 445 EYHISNT 452

RESULT 14

ABP71087
 ID ABP71087 standard; Protein; 1379 AA.

AC ABP71087;

DT 14-APR-2003 (first entry)

DE Human homologue of attractin/mahogany (HAM) polypeptide.

KW HAM; homologue of attractin/mahogany; immunosuppressive; cytostatic;
 KW antiinflammatory; cardiac; osteopathic; gene therapy; human.

OS Homo sapiens.

Key Location/Qualifiers

FT Peptide 1..60

FT Protein /note= "signal peptide"

FT Domain 61..1379

FT Domain /note= "mature protein"

FT Domain 63..90

FT Domain /note= "EGF-like domain"

FT Domain 211..244

FT Domain /note= "EGF-like domain"

FT Domain 261..280

FT Region /note= "EGF-like domain"

FT Region 581..612

FT Domain /note= "KELCH motif"

FT Domain 749..873

FT Domain /note= "laminin EGF-like domain"

FT Domain 1014..1056

FT Domain /note= "transmembrane domain"

WO200297120-A1.

PD 05-DEC-2002.

PF 23-MAY-2002; 2002WO-US16391.

PR 25-MAY-2001; 2001US-293608P.

PR 24-SEP-2001; 2001US-324626P.

PA (IMMUNEX CORP.

PI Anderson DW;

PI WPI: 2003-140486/13.

DR N-PSDB; ABZ58709.

DR New Homologue of Attractin/Mahogany (HAM) polypeptide, useful for

PT treating HAM-associated disorder consisting of inflammatory,

PT autoimmune, cell proliferative or cardiovascular disorders

XX Claim 3; Fig 1A-D; 89pp; English.

CC The invention relates to Homologue of Attractin/Mahogany (HAM)
 CC polypeptides and encoding polynucleotides. The HAM polypeptides can be
 CC expressed by standard recombinant methodology. The HAM polypeptides are
 CC useful for treating HAM-associated disorder consisting of inflammatory,
 CC autoimmune, graft-versus-host, neurological, myelination, cell
 CC proliferative, cardiovascular, haematologic, liver, metabolic, weight or
 CC bone disorder. The present sequence represents a human HAM polypeptide.
 XX

Sequence 1379 AA;

Query Match 1.9%; Score 8; DB 24; Length 1379;

Best Local Similarity 100.0%; Pred. No. 1.8e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 EYHISNT 71
 |||||
 446 EYHISNT 453

Db 446 EYHISNT 453

RESULT 15

AAU63540
 ID AAU63540 standard; Protein; 56 AA.

AC AAU63540;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #24436.

KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

PA (CORI-) CORIXA CORP.

PI Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhacila A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

PI WPI: 2001-616774/71.

DR N-PSDB; AAS59633.

DR Example 1; SEQ ID No 24735; 1069pp; English.

PT Propionibacterium acnes polypeptides and nucleic acids useful for

PT vaccinating against and diagnosing infections, especially useful for

PT treating acne vulgaris

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 56 AA;

Query Match 1.6%; Score 7; DB 22; Length 56;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 VFLSSR 34
DB 10 VFLSSR 16

RESULT 16
AAUS6058

ID AAUS6058 standard; Protein; 59 AA.

XX AAUS6058;

XX 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #16954.

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX MO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001MO-US12865.

XX 21-APR-2000; 2000US-199047P.

XX 02-JUN-2000; 2000US-208841P.

XX 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhactia A;

XX L' Maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

XX N-PSDB; AAS59573.

XX Example 1; SEQ ID No 17253; 1069bp; English.

CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to

CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 59 AA;

Query Match 1.6%; Score 7; DB 22; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 RRFOPAL 215
DB 4 RRFOPAL 10

RESULT 17
ABP34327

ID ABP34327 standard; Protein; 77 AA.

XX ABP34327;

XX 08-JUL-2002 (first entry)

XX Human ORF3300 protein, SEQ ID NO:6600.

KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; hematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; hemostatic;
KW chromolytic; tumour inhibition; bodily characteristics; fertility;
KW behaviour; cancer; proliferative disorder; neurological disorder;
KW cardiovascular disease; immune system disorder; organ transplantation;
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
KW vasotropic; antipruritic; antidiabetic; cyostatic; nootropic;
KW neuroprotective; antithrombotic; anticoagulant; thrombolytic;
KW cardiatic; hypotensive; antihypertoid; antiinflammatory; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide.

XX Homo sapiens.

XX MO200190366-A2.

XX 29-NOV-2001.

XX 24-MAY-2001; 2001MO-US17076.

XX 24-MAY-2000; 2000US-206690P.

XX (CURA-) CURAGEN CORP.

XX Leach MD, Shinkens RA;

XX WPI; 2002-106200/14.

XX N-PSDB; AEN78353.

XX Novel human polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and disorders related to organ
XX transplantation -

XX Claim 10; Page 1897; 2508bp; English.

CC Sequences ABP1028-ABP35561 represent 4534 novel human proteins
CC designated ORF (open reading frame) 1-4534, and sequences AEN75054-
CC AEN79587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX

CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, tumour inhibition activity,
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC and anti-infective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,
CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases.

CC SQ Sequence 77 AA;

CC Query Match

CC Best Local Similarity 1.6%; Score 7; DB 23; Length 77;

CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC QY 143 EKTGLGS 149

CC DB 66 EKTGLGS 72

CC RESULT 18

CC ID AAY08017 standard; Protein; 86 AA.

CC AC AAY08017;

CC DT 08-JUL-1999 (first entry)

CC DE Mouse partial TAP-1 protein.

CC LSP-1; signal peptide; leucocyte-specific protein-1; PA-I; TAP-1; cancer;
CC proliferin analogue I; thrombopoietin analogue protein 1; anticancer;
CC antiangiogenic; anticancer; anti-inflammatory; anti-thrombocytopenic;
CC anti-arthritic; signal transduction; inflammatory; disease; growth;
CC proliferation; differentiation; cell survival; angiogenesis; diagnosis;
CC haematopoietic stem cell; erythroid precursor; megakaryocytopoiesis;
CC thrombopoiesis; prognosis; treatment; chromosome mapping; tissue typing;
CC forensic; arthritis; thrombocytopenia; bone marrow transplant; infection;
CC intravascular coagulation; iron deficiency; HIV; murine.

CC OS Mus sp.

CC PN WO918243-A1.

CC PD 15-APR-1999.

CC PF 06-OCT-1998; 98WO-US21151.

CC XX 27-JAN-1998; 98US-0013447.

CC PR 06-OCT-1997; 97US-0061143.

CC PR 06-OCT-1997; 97US-0061149.

CC PR 08-JAN-1998; 98US-0004206.

PR 22-JAN-1998; 98US-0010674.

XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.

PI Gearing DP, McCarthy SA, Pan Y;

DR WPI; 1999-264042/22.

DR N-PSDB; AAX37562, AAX37563.

PT Signal peptide containing proteins that modulate cellular processes

PS Claim 1a; Fig 7; 124pp; English.

XX This invention describes the isolation of nucleic acids encoding the
XX signal-peptide-containing molecules leucocyte-specific protein-1 (LSP-1),
XX proliferin analog I (PA-I) and thrombopoietin analog protein 1 (TAP-1).
XX These proteins have antiangiogenic, anticancer, anti-inflammatory,
XX anti-arthritic and anti-thrombocytopenic activity. The products of the
XX invention and their modulators are involved in signal transduction,
XX inflammatory responses, growth, proliferation, differentiation and
XX survival of cells; angiogenesis; maturation of haematopoietic stem cells
XX and erythroid precursors megakaryocytopoiesis and thrombopoiesis.
XX Antibodies, or other binding agents, specific for the products of the
XX invention are useful for diagnosis, prognosis and monitoring of treatment
XX of diseases. Other uses include chromosome mapping, identification of
XX individuals (tissue typing) and in forensic studies. LSP-1, PA-I and
XX TAP-1 proteins and nucleic acids are modulators of cellular processes,
XX particularly they are used to treat or prevent diseases associated with
XX deregulation of angiogenesis, immune responses and haematopoiesis, e.g.
XX cancer, arthritis (and other inflammatory diseases), thrombocytopenia
XX (caused by cancer treatment, bone marrow transplant, human immune
XX deficiency virus infection etc.), intravascular coagulation, iron
XX deficiency etc.

CC SQ Sequence 86 AA;

CC Query Match

CC Best Local Similarity 1.6%; Score 7; DB 20; Length 86;

CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC QY 326 VEPLTVA 332

CC DB 34 VEPLTVA 40

CC RESULT 19

CC ID AAM87913 standard; Protein; 90 AA.

CC AC AAM87913;

CC DT 07-NOV-2001 (first entry)

CC DE Human immune/haematopoietic antigen SEQ ID NO:15506.

CC Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

CC cytosolic; gene therapy; vaccine; metastasis.

CC OS Homo sapiens.

CC PN WO200157182-A2.

CC PD 09-AUG-2001.

CC PF 17-JAN-2001; 2001WO-US01354.

CC XX 31-JAN-2000; 2000US-0179065.

CC PR 04-FEB-2000; 2000US-0180628.

CC PR 24-FEB-2000; 2000US-0184664.

CC PR 02-MAR-2000; 2000US-0186350.

CC PR 16-MAR-2000; 2000US-0189874.

CC PR 17-MAR-2000; 2000US-0190076.

CC PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214866.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216667.
PR 07-JUL-2000; 2000US-0216860.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 02-OCT-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
DR N-PSDB; AAK60694.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
PS Claim 11; SEQ ID NO 15506; 3071pp + Sequence Listing; English.
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)

CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (1) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (1) by expressing inactive proteins or to
CC supplement the patient's own production of (1). Additionally, (1)
CC polynucleotides may be used to produce the secreted (1), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (1) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.

SQ Sequence 90 AA;

Query Match 1.6%; Score 7; DB 22; Length 90;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 209 RRFOPAL 215
|||||
Db 74 RRFOPAL 80

RESULT 20
AAU36238
ID AAU36238 standard; Protein; 108 AA.

AC AAU36238;

DT 14-FEB-2002 (first entry)

DE Pseudomonas aeruginosa cellular proliferation protein #228.

KM Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.

OS Pseudomonas aeruginosa.

PN WO200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-265308P.

PA (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

XX N-PSDB; AAS54097.

XX

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 108 AA;

Query Match 1.6%; Score 7; DB 22; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 151 AGLVSVV 157
|||||
Db 38 AGLVSVV 44

RESULT 21
ABJ18790
ID ABJ18790 standard; Protein; 108 AA.

AC ABJ18790;

DT 27-FEB-2003 (first entry)

DE Pseudomonas aeruginosa biofilm formation-related protein #54.

KM Biofilm formation modulation; biofilm-associated disease;
KW cystic fibrosis; AIDS; middle ear infection; acne; periodontal disease;
KW catheter-associated infection; medical device-associated infection.

OS Pseudomonas aeruginosa.

PN WO200285295-A2.

PD 31-OCT-2002.

PF 19-APR-2002; 2002WO-US12532.

PR 20-APR-2001; 2001US-285190P.

PR 24-OCT-2001; 2001US-344142P.

PR (IOWA) UNIV IOWA RES FOUND.

PR (HARD) HARVARD COLLEGE.

PA Whiteley M, Bangera MG, Lory S, Greenberg EP;

PI WPI; 2003-075601/07.

PI N-PSDB; ABT14612.

XX

XX Claim 1; Page 132; 154pp; English.

XX The invention comprises a method for identifying a compound capable of
XX modulating biofilm formation by bacteria. The method of the invention is
XX useful for identifying a compound capable of modulating biofilm formation
XX by bacteria or modulating bacterial antibiotic resistance. The method of
XX the invention is also useful for diagnosing and treating a subject
XX (especially an immunocompromised human) that is afflicted with a biofilm-
XX associated disease or disorder, such as: cystic fibrosis; AIDS; middle
XX ear infections; acne; periodontal disease; catheter-associated

CC infections; and medical device-associated infections. The present amino
CC acid sequence represents a protein that is used in the invention.
XX

SO Sequence 108 AA;

Query Match 1.6%; Score 7; DB 24; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 AGLVSVV 157
Db 38 AGLVSVV 44

RESULT 22

AAG40712
ID AAG40712 standard; Protein; 109 AA.

XX AAG40712;

DT 18-OCT-2000 (first entry)

DE Zea mays protein fragment SEQ ID NO: 50552.

XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence; corn.

OS Zea mays subsp. mays.

XX EPI033405-A2.

XX PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135533.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 26-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.6%; Score 7; DB 21; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 233 ELKLEIE 239
Db 77 ELKLEIE 83

RESULT 23
AAM34293
ID AAM34293 standard; Protein; 115 AA.

XX AAM34293;

XX 17-OCT-2001 (first entry)

XX Peptide #8330 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

XX Claim 27; SEQ ID No 34562; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs;

XX see AAI31315-AI57546). The present sequence is a peptide encoded by one

XX such probe. The probes are useful for producing a microarray for

XX predicting, measuring and displaying gene expression in samples derived

XX from human placenta. The probes are useful for antenatal diagnosis of

XX human genetic disorders.

XX Sequence 115 AA;

XX SQ

Query Match 1.6%; Score 7; DB 22; Length 115;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 264 GGSSETPK 270
Db 37 GGSSETPK 43

RESULT 24

AAAG41932
ID AAAG41932 standard; Protein; 124 AA.

XX AAAG41932;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 52232.
XX
XX Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX
PD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135829.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.

PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155559.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 06-OCT-1999; 99US-0157753.
 PR 07-OCT-1999; 99US-0157865.
 PR 08-OCT-1999; 99US-0158029.
 PR 12-OCT-1999; 99US-0158232.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159684.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.6%; Score 7; DB 21; Length 124;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 VATSLIS 163
 |||||
 Db 76 VATSLIS 82

RESULT 25
 AAY36977
 ID AAY36977 standard; Protein; 132 AA.

AC AAY36977;
 XX

DT 07-OCT-1999 (first entry)

XX Amino acid sequence of a Chlamydia trachomatis protein.

DE Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KW paratrachoma; inclusion conjunctivitis; genital disease; perithenaritis;
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KW bartonellosis; pneumopathy; venereal lymphogranulomatosis.
 XX

OS Chlamydia trachomatis.
 XX
 PN WO9928475-A2.
 XX
 PD 10-JUN-1999.
 XX
 PF 27-NOV-1999; 98WO-IB01939.
 XX
 PR 04-NOV-1998; 98US-0107077.
 PR 28-NOV-1997; 97FR-0015041.
 PR 17-DEC-1997; 97FR-0016034.
 XX
 PA (GENSET) GENSET.
 XX
 PI Griffiths R;
 XX
 DR WPI, 1999-371125/31.
 XX
 PT Genome sequence of Chlamydia trachomatis
 XX
 PS Disclosure; Page 816, 1755pp; English.
 XX
 CC AAY6754-137949 are encoded by open reading frames (ORFs) of the genome
 CC of Chlamydia trachomatis (see AAY01425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nonendemic trachoma,
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC perithenaritis, bartonellosis; pneumopathy in breast feeding infants;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.
 XX
 SQ Sequence 132 AA;

Query Match 1.6%; Score 7; DB 20; Length 132;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 LGSSAGL 153
 |||||
 Db 67 LGSSAGL 73

RESULT 26
 ABB59022
 ID ABB59022 standard; Protein; 137 AA.

AC ABB59022;
 XX

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 3856.

KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PERK) PERK CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;
 XX

DR WPI; 2001-656860/75.
DR N-PSDB; ABL03125.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 3858; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB161840-AB16175) and the encoded proteins
CC (ABBS7737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 137 AA;
XX
Query Match 1.6%; Score 7; DB 22; Length 137;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 12 FLAGGYL 18
DB 22 FLAGGYL 28
XX
RESULT 27
ID AAM81086
AC AAM81086 standard; Protein; 146 AA.
XX
AC AAM81086;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human haematological malignancy-related antigen #784.
XX
KW Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;
KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX
OS Homo sapiens.
XX
PN WO200164886-A2.
XX
PD 07-SEP-2001.
XX
PF 01-MAR-2001; 2001WO-US07272.
XX
PR 01-MAR-2000; 2000US-0186126.
PR 17-MAR-2000; 2000US-0190479.
PR 27-APR-2000; 2000US-0200545.
PR 28-APR-2000; 2000US-0200303.
PR 28-APR-2000; 2000US-0200779.
PR 01-MAY-2000; 2000US-0200999.
PR 04-MAY-2000; 2000US-0202084.
PR 22-MAY-2000; 2000US-0206201.
PR 14-JUL-2000; 2000US-0218950.
PR 03-AUG-2000; 2000US-0222903.
PR 04-AUG-2000; 2000US-0223416.
PR 07-AUG-2000; 2000US-0223378.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Algate PA, Mannion J;
XX
DR WPI; 2001-514842/56.
XX
PT

PT Compositions and methods for the detection of hematological
PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX
PS Claim 1, Page 828, 1252pp; English.
XX
CC The present invention relates to compositions and methods for the
CC detection, diagnosis and therapy of haematological malignancies. The
CC present sequence is the protein sequence of a human haematological
CC malignancy related antigen. The methods of the present invention comprise
CC detecting the presence of haematological malignancy related antigen(s) in
CC a sample obtained from the patient (an increased level of the
CC polypeptide, compared to an unaffected individual, is indicative of an
CC increased risk). Haematological malignancies which can be treated using
CC the present invention are chronic lymphocytic leukaemia, lymphoma,
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
CC cell non-Hodgkin's lymphoma.
XX
SQ Sequence 146 AA;
XX
Query Match 1.6%; Score 7; DB 22; Length 146;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 233 ELKKLIE 239
DB 56 ELKKLIE 62
XX
RESULT 28
ID AAM81536
AC AAM81536 standard; Protein; 146 AA.
XX
AC AAM81536;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human haematological malignancy-related antigen #1234.
XX
KW Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;
KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX
OS Homo sapiens.
XX
PN WO200164886-A2.
XX
PD 07-SEP-2001.
XX
PF 01-MAR-2001; 2001WO-US07272.
XX
PR 01-MAR-2000; 2000US-0186126.
PR 17-MAR-2000; 2000US-0190479.
PR 27-APR-2000; 2000US-0200545.
PR 28-APR-2000; 2000US-0200303.
PR 28-APR-2000; 2000US-0200779.
PR 01-MAY-2000; 2000US-0200999.
PR 04-MAY-2000; 2000US-0202084.
PR 22-MAY-2000; 2000US-0206201.
PR 14-JUL-2000; 2000US-0218950.
PR 03-AUG-2000; 2000US-0222903.
PR 04-AUG-2000; 2000US-0223416.
PR 07-AUG-2000; 2000US-0223378.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Algate PA, Mannion J;
XX
DR WPI; 2001-514842/56.
XX
PT Compositions and methods for the detection of hematological
PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -

XX Claim 1, Page 1003; 1252pp; English.
 PS
 CC The present invention relates to compositions and methods for the
 CC detection, diagnosis and therapy of hematological malignancies. The
 CC present sequence is the protein sequence of a human haematological
 CC malignancy related antigen. The methods of the present invention comprise
 CC detecting the presence of hematological malignancy related antigen(s) in
 CC a sample obtained from the patient (an increased level of the
 CC polypeptide, compared to an unaffected individual, is indicative of an
 CC increased risk). Hematological malignancies which can be treated using
 CC the present invention are chronic lymphocytic leukaemia, lymphoma,
 CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
 CC cell non-Hodgkin's lymphoma.
 CC
 SQ Sequence 146 AA;
 Query Match 1.6%; Score 7; DB 22; Length 146;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 233 ELKKE 239
 |||||
 56 ELKKE 62
 Db
 RESULT 29
 ABU61012
 ID ABU61012 standard; Protein; 149 AA.
 AC ABU61012;
 XX
 DT 08-MAY-2003 (first entry)
 XX
 DE Lung specific protein (LSP) #115.
 XX
 KW Human; gene therapy; vaccine; lung specific antigen; cancer diagnosis;
 KW cancer monitoring; cancer staging; cancer imaging; lung cancer;
 KW non-cancerous diseases of the lung; transgenic animal.
 OS Homo sapiens.
 XX
 PN WO200266633-A2.
 XX
 PD 06-SEP-2002.
 XX
 PF 21-NOV-2001; 2001WO-US43612.
 XX
 PR 22-NOV-2000; 2000US-252500P.
 XX
 PA (DIAD-) DIADEXUS INC.
 XX
 PI Macina RA, Recipon H, Chen S, Sun Y, Liu C,
 XX
 DR WPI; 2002-713376/77.
 XX
 PT New isolated human nucleic acid molecule and polypeptide, useful for
 PT identifying, diagnosing, monitoring, staging, imaging and treating lung
 PT cancer and non-cancerous diseases of the lung
 XX
 PS Claim 11; Page 386; 389pp; English.
 XX
 CC The invention describes an isolated human nucleic acid (I) encoding any
 CC of 120 10-1533 residue amino acid sequences (S1), given in the
 CC specification, comprising any of 164 179-12421 base pair sequences (S2),
 CC given in the specification. The methods and compositions of the present
 CC invention are useful for identifying, diagnosing, monitoring, staging,
 CC imaging and treating lung cancer and non-cancerous diseases of the lung.
 CC They are also used for identifying lung tissue, monitoring and
 CC identifying and/or designing antagonists of the polypeptide of the
 CC invention, gene therapy, production of transgenic animals and production
 CC of engineered lung tissue for treatment and research. This is the amino
 CC acid sequence of a lung specific nucleic acid.

XX Sequence 149 AA;
 SQ
 Query Match 1.6%; Score 7; DB 23; Length 149;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 369 CVGQVVP 375
 |||||
 124 CVGQVVP 130
 Db
 RESULT 30
 ABP64574
 ID ABP64574 standard; Protein; 157 AA.
 AC ABP64574;
 XX
 DT 04-NOV-2002 (first entry)
 XX
 DE Human ORF944.
 XX
 KW Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnerary;
 KW Antiinflammatory; gene therapy; human; ORFX; atherogenic; platelet;
 KW human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;
 KW cancer; cardiovascular disease; allergy; autoimmune disease;
 KW wound healing; blood coagulation disorder; inflammatory disorder.
 OS Homo sapiens.
 XX
 PN US2002082206-A1.
 XX
 DT 27-JUN-2002.
 XX
 PF 30-MAY-2001; 2001US-0867550.
 XX
 PR 30-MAY-2000; 2000US-208427P.
 XX
 PA (LEACH/) LEACH M D.
 PA (MEHR/) MEHRABAN F.
 PA (CONL/) CONLEY P B.
 PA (TOPP/) TOPPER J N.
 PA (LAWD/) LAW D.
 XX
 PI Leach MD, Mehraban F, Conley PB, Topper JN, Law D;
 XX
 DR WPI; 2002-626554/67.
 DR N-PSDB; ABO99137.
 XX
 PT New polypeptide designated ORFX are present in human atherogenic cells
 PT and are useful to prevent and treat ORFX-associated disorders including
 PT cancer, allergy, wound healing or autoimmune, cardiovascular or
 PT inflammatory disease
 XX
 PS Claim 10; SEQ ID 1888; 78pp; English.
 XX
 CC The present invention relates to novel human ORFX polypeptides and their
 CC coding sequences (ABP63631-ABP64681 and ABO98194-ABO99267). The sequences
 CC were discovered in human atherogenic cells, in particular in platelets
 CC and human umbilical vein endothelial cells (HUVEC) and are expressed in
 CC many other tissues as well. Atherogenic cells are cells which have the
 CC potential to develop atherosclerotic plaques. The ORFX polypeptides and
 CC nucleic acids are useful for treating or preventing a pathological
 CC condition associated with an ORFX-associated disorder, e.g. cancer,
 CC cardiovascular disease, allergy, autoimmune disease, wound healing, blood
 CC coagulation disorders or inflammatory disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site at seqdata.uspto.gov/sequence.html?docid=2002082206:
 XX
 SQ Sequence 157 AA;
 Query Match 1.6%; Score 7; DB 23; Length 157;

Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 RRFOPAL 215
| | | | |
DB 110 RRFOPAL 116

RESULT 31
AA84688
ID AA84688 standard; Protein; 176 AA.

AC AA84688;

DT 08-AUG-2000 (first entry)

XX Amino acid sequence of AG005, a phosphatidylinositol-4 kinase protein.

XX GTP-binding protein; AG004; GTPase activating protein; fungal growth;
KW phosphatidylinositol-4 kinase; cytokinesis gene; fungal development;
KW fungicidal; filamentous fungi; plant pathogen; Septoria tritici;
KW Stagnospora nodorum; Magnaporthe grisea; human pathogen;
KW Candida albicans; Aspergillus fumigatus.

XX Ashbya gossypii.

PN W0200022133-A1.

PD 20-APR-2000.

PF 06-OCT-1999; 99MO-EP07501.

PR 08-OCT-1998; 98US-0168804.

PA (NOVS) NOVARTIS-ERFINDUNGEN VERM GBS MBH.

PI Gaffney TD, Wendland J, Dietrich F, Philippsen P, Goff SA;

DR WPI; 2000-317988/27.

DR N-PSDB; AAA14505.

PT Ashbya gossypii nucleotide and protein sequences useful for identifying
XX fungal growth inhibitors

PS Claim 9; Page 76-77; 93pp; English.

XX The present sequence represents a putative phosphatidylinositol-4 kinase
CC protein, designated AG005. The specification also describes putative
CC GTP binding protein genes, putative GTPase activating protein genes,
CC and a putative cytokinesis gene. These genes are essential for
CC fungal growth and development. The proteins can be used in methods
CC to identify compounds that have fungicidal activity. Compounds
CC with fungicidal activity can be used for suppressing fungal growth,
CC especially of filamentous fungi. Fungi that can be suppressed include
CC plant pathogens (e.g. Septoria tritici, Stagnospora nodorum, and
CC Magnaporthe grisea) and human pathogens (e.g. Candida albicans, and
CC Aspergillus fumigatus).

SQ Sequence 176 AA;

Query Match 1.6%; Score 7; DB 21; Length 176;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 328 PLTVAIK 334
| | | | |
DB 55 PLTVAIK 61

RESULT 32
AAE11054
ID AAE11054 standard; Protein; 176 AA.

XX AAE11054;
AC
XX 18-DEC-2001 (first entry)

DT Ashbya gossypii phosphatidylinositol-4 kinase protein, AG005.

XX Ashbya gossypii phosphatidylinositol-4 kinase protein; AG005; fungicide;
KW chromosome mapping; fungal growth; chromosome identification.

XX Ashbya gossypii.

PN US6291660-B1.

PD 18-SEP-2001.

PF 08-OCT-1999; 99US-0415522.

PR 08-OCT-1998; 98US-172224P.

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

PI Gaffney TD, Wendland J, Dietrich F, Philippsen P, Goff SA;

DR WPI; 2001-595520/67.

DR N-PSDB; AAD18407.

PT Fungal genes required for normal growth and development -

PS Example 1; Column 55-56; 33pp; English.

XX The invention relates to Ashbya gossypii specific genes (AG001-AG006) and
CC their corresponding protein molecules. Ashbya gossypii specific genes are
CC essential for fungal growth and development. The invention is useful in
CC screening assays to identify inhibitors that are used as fungicides to
CC suppress the growth of pathogenic fungi. The invention is also useful for
CC identifying biosynthetic genes which are used as selection markers, to
CC isolate promoters and terminators for application in a homologous as well
CC as heterologous context, to find putative centromere containing clones,
CC chromosome mapping, chromosome identification and general information
CC about chromosome organization. The present sequence is Ashbya gossypii
CC phosphatidylinositol-4 kinase protein, AG005.

SQ Sequence 176 AA;

Query Match 1.6%; Score 7; DB 22; Length 176;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 328 PLTVAIK 334
| | | | |
DB 55 PLTVAIK 61

RESULT 33
AAG40711
ID AAG40711 standard; Protein; 178 AA.

AC AAG40711;

DT 18-OCT-2000 (first entry)

DE Zea mays protein fragment SEQ ID NO: 50551.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence; corn.

OS Zea mays subsp. mays.

PN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.
 PR 25-FEB-1999; 99US-0121825.
 XX 25-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138647.
 PR 14-JUN-1999; 99US-0138119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145912.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145914.
 PR 27-JUL-1999; 99US-0145915.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148684.
 PR 13-AUG-1999; 99US-0148685.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149910.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.

PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158366.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.6%; Score 7; DB 21; Length 178;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ELKKLE 239
 DB 146 ELKKLE 152

RESULT 34
 ID ABP64892 standard; Protein; 179 AA.

AC ABP64892;

DT 25-FEB-2003 (first entry)

DE Human protein SEQ ID 552.

XX Human; expressed sequence tag; EST;
 KW haematopoietic disorder; central nervous system disease; viral infection;
 KW peripheral nervous system disease; non-healing wound; infectious disease;
 KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;
 KW fungal infection; autoimmune disorder; coagulation disorder; neutropenic;
 KW anti-infective; anti-inflammatory; immunosuppressive; neuroprotective;
 KW cytotoxic; haemostatic; virucide; antibacterial; fungicide;
 KW immunostimulant; cerebroprotective.

OS Homo sapiens.

PN WO200259260-A2.

XX 01-AUG-2002.

PF 16-NOV-2001; 2001WO-US42950.

XX 17-NOV-2000; 2000US-0714936.
 PR (HYSE-) HYSEQ INC.
 XX
 PA Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;
 PI Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;
 XX WPI; 2002-590824/63.
 DR N-PSDB; ABQ99478.

PT New isolated polynucleotide, useful in research, diagnostic or
 PT therapeutic methods, e.g. preventing or treating disorders involving
 PT aberrant protein expression or biological activity -

PS Claim 20; SEQ ID 552; 394pp; English.

CC The present invention relates to novel human coding sequences
 CC (ABQ99268-ABQ99608) and proteins (ABP64892-ABP65022). The sequences are
 CC useful in therapeutic, diagnostic and research methods. The
 CC polynucleotides may be used in the field of molecular biology as
 CC hybridisation probes, primers for PCR, for chromosome and gene mapping,
 CC for the recombinant production of protein, or in generation of anti-sense
 CC DNA or RNA. The polynucleotides are useful in diagnostics as expressed
 CC sequence tags (ESTs) for identifying expressed genes or for physical
 CC mapping of the human genome. The proteins may be used as molecular weight
 CC markers, or as nutritional sources or supplements. The proteins may be
 CC used to maintain and expand cell population in a totipotent or
 CC pluripotent state useful for re-engineering damaged or diseased
 CC tissues, transplantation, manufacture of bio-pharmaceuticals or the
 CC development of bio-sensors. The polynucleotides and proteins are useful
 CC for preventing, treating or ameliorating disorders involving aberrant
 CC protein expression or biological activity, e.g. haematopoietic disorders,
 CC central/peripheral nervous system diseases, mechanical and traumatic
 CC disorders, non-healing wounds, immune deficiencies and disorders,
 CC infectious diseases caused by viral, bacterial or fungal infection,
 CC autoimmune disorders, allergic reactions and conditions, coagulation
 CC disorders, or cancer. The polynucleotide sequences of the invention were
 CC assembled from ESTs isolated mainly by sequencing by hybridisation, and
 CC in some cases, sequences obtained from one or more public databases.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 179 AA;

Query Match 1.6%; Score 7; DB 23; Length 179;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 EATIFIV 92
 DB 86 EATIFIV 92

RESULT 35

ID AAG34111 standard; Protein; 183 AA.

AC AAG34111;

DT 18-OCT-2000 (first entry)

DE Zea mays protein fragment SEQ ID NO: 41455.

XX Protein identification; signal transduction pathway; metabolic pathway;

XX hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence; corn.

OS Zea mays subsp. mays.

PN EP1033405-A2.

PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0155648.
 PR 29-SEP-1999; 99US-0155656.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0158293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.6%; Score 7; DB 21; Length 183;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 APGKAF 13
 |||||
 DB 173 APGKAF 179

RESULT 36
 AAE13841
 ID AAE13841 standard; Protein; 184 AA.
 XX
 AC AAE13841;
 XX
 DT 26-FEB-2002 (first entry)
 XX
 DE Human lung tumour-specific protein SCC2-60.
 XX
 KW Human; lung tumour protein; immunostimulant; cytosratic; gene therapy;
 KW antisense-therapy; vaccine; immune response; lung cancer; SCC2-60.
 XX
 OS Homo sapiens.
 XX
 PN WO200172295-A2.
 XX
 PD 04-OCT-2001.
 XX
 PF 28-MAR-2001; 2001WO-US09991.
 XX
 PR 29-MAR-2000; 2000US-0538037.
 PR 05-JUN-2000; 2000US-0588937.
 PR 18-AUG-2000; 2000US-0640878.
 PR 22-SEP-2000; 2000US-234517P.

PR 01-NOV-2000; 2000US-0704512.
 PR 14-DEC-2000; 2000US-0738973.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;
 PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannon J, Kalos MD;
 XX
 DR WPI; 2001-639201/73.
 DR N-PSDB; AAD23453.
 XX
 PT New human lung-specific polynucleotides and polypeptides for the
 PT diagnosis and treatment of disease e.g. lung cancer -
 XX
 PS Disclosure; Page 319; 378pp; English.
 XX
 CC The invention relates to isolated lung tumour-specific proteins and
 CC their corresponding cDNA molecules. Lung tumour-specific proteins and
 CC their antigen-presenting cells are useful for stimulating and/or
 CC expanding T cells specific for a tumour protein, and for inhibiting
 CC the development of cancer. The invention also relates to a composition
 CC useful for stimulating an immune response, and for treating cancer. The
 CC lung tumour specific oligonucleotide is useful in gene therapy and for
 CC diagnosis, detection and treatment of lung cancer. The present sequence
 CC is human lung tumour-specific protein.
 XX
 SQ Sequence 184 AA;

Query Match 1.6%; Score 7; DB 22; Length 184;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 EATIRIV 92
 |||||
 DB 91 EATIRIV 97

RESULT 37
 AB05936
 ID AB05936 standard; Protein; 186 AA.
 XX
 AC AB05936;
 XX
 DT 08-APR-2003 (first entry)
 XX
 DE M. tuberculosis and M. leprae marker protein #587.
 XX
 KW Mycobacterioses; survival; virulence; protective antigen; vaccine;
 KW mycobacterial disease; tuberculosis; leprosy.
 XX
 OS Mycobacterium tuberculosis.
 OS Mycobacterium leprae.
 XX
 PN WO200274903-A2.
 XX
 PD 26-SEP-2002.
 XX
 PF 22-FEB-2002; 2002WO-IB01973.
 XX
 PR 22-FEB-2001; 2001US-270123P.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Cole S;
 XX
 DR WPI; 2002-759885/82.
 XX
 PT Identifying and selecting genes for survival or virulence of
 PT mycobacteria by a comparative genomic analysis of the sequences of
 PT Mycobacterium tuberculosis and M. leprae -
 XX
 PS Claim 17; Page 804; 874pp; English.
 XX

CC This invention relates to a novel method for identifying essential genes
CC for survival or virulence of mycobacteria species. The method comprises
CC aligning the genomic sequence of a first mycobacterium species on a
CC genomic sequence of a second mycobacterium species and selecting a
CC polynucleotide sequence that is highly conserved in both genomes with no
CC counterparts in other bacterial genomic sequences and that corresponds
CC to an essential gene for the survival or virulence of mycobacterium
CC species. The method of the invention is useful for detecting M.
CC tuberculosis or M. leprae infection. The method reduces the number of
CC potential new targets and protective antigens for new drugs and vaccine
CC compositions to treat and prevent mycobacterial diseases, particularly
CC tuberculosis and leprosy. The present sequence represents a marker
CC protein from Mycobacterium tuberculosis and Mycobacterium leprae
CC identified using the method of the invention.

SQ Sequence 186 AA;

Query Match 1.6%; Score 7; DB 23; Length 186;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 SVVATSL 161
Db 150 SVVATSL 156

RESULT 38

ABG68071 standard; Protein; 189 AA.

AC ABG68071;

DT 24-SEP-2002 (first entry)

DE Human interferon alpha (IFN alpha) #13.

KW Interferon; immune system-related disorder; viral infection; cancer;
KW parasitic infection; bacterial infection; autoimmune disease;
KW multiple sclerosis; lymphoma; allergy; hairy cell leukaemia; hepatitis C;
KW Kaposi's sarcoma; chronic myelogenous leukaemia; multiple myeloma;
KW basal cell carcinoma; malignant melanoma; ovarian cancer; hepatitis D;
KW cutaneous T cell lymphoma; anti-viral therapy; acquired immune disorder;
KW chronic hepatitis B; papilloma virus infection; vaccine adjuvants;
KW multidrug-resistant pulmonary tuberculosis; rabies; feline panleukopenia;
KW feline leukaemia virus infections; feline infectious peritonitis;
KW inflammatory airway disease; human.

OS Homo sapiens.

PN WO20023627-A2.

PD 10-MAY-2002.

PF 05-NOV-2001; 2001WO-US47226.

PR 03-NOV-2000; 2000US-245754P.

PR 03-NOV-2000; 2000US-246234P.

PA (PBLB-) PBL BIOMEDICAL LAB.

XX Pestka S;

XX MPI; 2002-519235/55.

XX N-PSDB; ABK96751.

PT Novel isolated interferon polypeptide and polynucleotides encoding the
PT polypeptide, useful for treating an immune system-related disorder e.g.
PT viral, parasitic or bacterial infections or allergy, in a non-human
PT animal

PS Claim 1; Page 91; 97bp; English.

CC The invention describes an isolated interferon polypeptide (1).

CC (1) is useful for treating an immune system-related disorder, such as
CC viral infection, parasitic infection, bacterial infection, cancer,
CC autoimmune disease, multiple sclerosis, lymphoma, or allergy in a
CC patient, preferably a non-human animal, cancer, hairy cell leukaemia,
CC Kaposi's sarcoma, chronic myelogenous leukaemia, multiple myeloma, basal
CC cell carcinoma, malignant melanoma, ovarian cancer and cutaneous T cell
CC lymphoma. (1) can also be used for anti-viral therapy e.g. in the
CC treatment of acquired immune disorders, e.g. chronic hepatitis B,
CC hepatitis C, hepatitis D, papilloma virus infections, etc. (1) can be
CC used as part of an immunotherapy protocol, or as vaccine adjuvants. (1)
CC is also useful for treating bacterial infections e.g. multidrug-resistant
CC pulmonary tuberculosis. Most preferably (1) is useful for treating cats
CC as part of veterinarian protocols, and thus are useful for treating viral
CC infections (e.g. feline leukaemia virus infections), feline
CC panleukopenia, feline infectious peritonitis, rabies, inflammatory airway
CC disease, in cats. (1) is also useful for treating dogs or other household
CC pets, and other farm animals. This is the amino acid sequence of a human
CC interferon-alpha described in the invention.

SQ Sequence 189 AA;

Query Match 1.6%; Score 7; DB 23; Length 189;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 EKPTTEL 234
Db 106 EKPTTEL 112

RESULT 39

ABG68076 standard; Protein; 189 AA.

AC ABG68076;

DT 24-SEP-2002 (first entry)

DE Human interferon alpha (IFN alpha) #18.

KW Interferon; immune system-related disorder; viral infection; cancer;
KW parasitic infection; bacterial infection; autoimmune disease;
KW multiple sclerosis; lymphoma; allergy; hairy cell leukaemia; hepatitis C;
KW Kaposi's sarcoma; chronic myelogenous leukaemia; multiple myeloma;
KW basal cell carcinoma; malignant melanoma; ovarian cancer; hepatitis D;
KW cutaneous T cell lymphoma; anti-viral therapy; acquired immune disorder;
KW chronic hepatitis B; papilloma virus infection; vaccine adjuvants;
KW multidrug-resistant pulmonary tuberculosis; rabies; feline panleukopenia;
KW feline leukaemia virus infections; feline infectious peritonitis;
KW inflammatory airway disease; human.

OS Homo sapiens.

PN WO20023627-A2.

PD 10-MAY-2002.

PF 05-NOV-2001; 2001WO-US47226.

PR 03-NOV-2000; 2000US-245754P.

PR 03-NOV-2000; 2000US-246234P.

PA (PBLB-) PBL BIOMEDICAL LAB.

XX Pestka S;

XX MPI; 2002-519235/55.

XX N-PSDB; ABK96756.

PT Novel isolated interferon polypeptide and polynucleotides encoding the
PT polypeptide, useful for treating an immune system-related disorder e.g.
PT viral, parasitic or bacterial infections or allergy, in a non-human
PT animal

XX Claim 1; Page 94; 97pp; English.
 PS
 CC The invention describes an isolated interferon polypeptide (I).
 CC (I) is useful for treating an immune system-related disorder, such as
 CC viral infection, parasitic infection, bacterial infection, cancer,
 CC autoimmune disease, multiple sclerosis, lymphoma, or allergy in a
 CC patient, preferably a non-human animal, cancer, hairy cell leukaemia,
 CC Kaposi's sarcoma, chronic myelogenous leukaemia, multiple myeloma, basal
 CC cell carcinoma, malignant melanoma, ovarian cancer and cutaneous T cell
 CC lymphoma. (I) can also be used for anti-viral therapy e.g. in the
 CC treatment of acquired immune disorders, e.g. chronic hepatitis B,
 CC hepatitis C, hepatitis D, papilloma virus infections, etc. (I) can be
 CC used as part of an immunotherapy protocol, or as vaccine adjuvants. (I)
 CC is also useful for treating bacterial infections e.g. multidrug-resistant
 CC pulmonary tuberculosis. Most preferably (I) is useful for treating viral
 CC as part of a veterinarian protocol, and thus are useful for treating viral
 CC infections (e.g. feline leukaemia virus infections), feline
 CC panleukopenia, feline infectious peritonitis, rabies, inflammatory airway
 CC disease, in cats. (I) is also useful for treating dogs or other household
 CC pets, and other farm animals. This is the amino acid sequence of a human
 CC interferon-alpha described in the invention.
 CC
 SQ Sequence 189 AA;
 Query Match 1.6%; Score 7; DB 23; Length 189;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 228 EKPTEL 234
 |||||
 DB 106 EKPTEL 112
 RESULT 40
 ID ABP42703
 ID ABP42703 standard; Protein; 193 AA.
 AC ABP42703;
 XX
 DT 22-AUG-2002 (first entry)
 XX
 DE Human ovarian antigen HOPK47, SEQ ID NO:3835.
 XX
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive.
 KW
 OS Homo sapiens.
 XX
 PN WO200200677-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 07-JUN-2001; 2001WO-US18569.
 XX
 PR 07-JUN-2000; 2000US-209467P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Birze CE, Rosen CA;
 XX
 DR WPI: 2002-147678/19.
 DR N-PSDB; AB055780.
 XX
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,

PT useful in the prevention, treatment and diagnosis of cancer (e.g.
 PT ovarian cancer), immune disorders, cardiovascular disorders and
 PT neurological diseases -
 XX
 PS Claim 11; SEQ ID NO 3835; 2922pp; English.
 XX
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (AB054131-AB056305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 193 AA;
 Query Match 1.6%; Score 7; DB 23; Length 193;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 AFSAPGK 10
 |||||
 DB 2 AFSAPGK 8
 Search completed: August 8, 2003, 09:50:21
 Job time : 85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2003, 09:39:52 ; Search time 52 Seconds

(without alignments)
986.620 Million cell updates/sec

Title: US-10-069-062-7

Perfect score: 2230

Sequence: 1 MSKAPSGAKFLAGYIVL.....DLDEQTEGLVEKPEYDYL 432

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 segs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PTC_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PTCUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	809.5	36.3	451	US-09-909-745-24	Sequence 24, Appl
2	671	30.1	451	US-10-128-714-3406	Sequence 3406, Ap
3	655.5	29.4	484	US-10-128-714-8406	Sequence 8406, Ap
4	475	21.3	505	US-09-988-863A-2	Sequence 2, Appl
5	445.5	20.0	503	US-10-036-959B-12	Sequence 12, Appl
6	209.5	9.4	368	US-09-815-242-10656	Sequence 10656, A
7	181.5	8.1	179	US-09-909-745-22	Sequence 22, Appl
8	173	7.8	345	US-09-815-242-5854	Sequence 5854, Ap
9	173	7.8	358	US-09-815-242-12866	Sequence 12866, A
10	123	5.5	378	US-09-909-745-23	Sequence 23, Appl
11	122	5.5	335	US-09-815-242-13670	Sequence 13670, A
12	107.5	4.8	743	US-10-087-464-53	Sequence 53, Appl
13	105	4.7	54	US-09-909-745-16	Sequence 16, Appl
14	105	4.7	54	US-09-909-745-18	Sequence 18, Appl
15	104.5	4.7	293	US-09-902-525-42	Sequence 42, Appl

16	101.5	4.6	859	US-09-978-522-3	Sequence 3, Appl
17	100	4.5	862	US-09-978-522-1	Sequence 1, Appl
18	100	4.5	1174	US-10-205-841-40	Sequence 40, Appl
19	99.5	4.5	1734	US-09-862-027-81	Sequence 81, Appl
20	99	4.4	394	US-09-862-027-19	Sequence 19, Appl
21	98	4.4	1272	US-09-769-097-2	Sequence 2, Appl
22	98	4.4	1272	US-09-769-097-4	Sequence 4, Appl
23	98	4.4	1457	US-09-772-316-2	Sequence 2, Appl
24	96.5	4.3	930	US-09-815-028A-17	Sequence 17, Appl
25	96.5	4.3	496	US-09-815-242-5644	Sequence 5644, Ap
26	96.5	4.3	496	US-09-815-242-12657	Sequence 12657, A
27	96	4.3	3907	US-10-171-311-2	Sequence 2, Appl
28	96	4.3	3925	US-10-171-311-6	Sequence 6, Appl
29	95.5	4.3	930	US-09-815-242-10779	Sequence 10779, A
30	95.5	4.3	3899	US-10-171-311-4	Sequence 4, Appl
31	95.5	4.3	3917	US-10-171-311-8	Sequence 8, Appl
32	95	4.3	382	US-09-800-729-206	Sequence 206, App
33	95	4.3	382	US-09-987-107-37	Sequence 37, Appl
34	95	4.3	386	US-10-036-959B-11	Sequence 11, Appl
35	95	4.3	478	US-10-156-761-10884	Sequence 10884, A
36	94.5	4.2	590	US-09-925-300-1219	Sequence 1219, Ap
37	94	4.2	901	US-09-828-062-8	Sequence 8, Appl
38	94	4.2	1616	US-09-820-843A-16	Sequence 16, Appl
39	93.5	4.2	396	US-09-987-107-33	Sequence 33, Appl
40	93.5	4.2	396	US-09-802-640-16	Sequence 16, Appl
41	93.5	4.2	1169	US-09-815-242-13448	Sequence 13448, A
42	93.5	4.2	1838	US-09-879-248-8	Sequence 8, Appl
43	93	4.2	231	US-09-842-758-107	Sequence 107, App
44	93	4.2	333	US-09-764-864-1010	Sequence 1010, App
45	93	4.2	639	US-09-934-455-426	Sequence 426, App

ALIGNMENTS

RESULT 1
US-09-909-745-24
; Sequence 24, Application US/09909745
; Patent No. US20020119546A1
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Omolayo O.
; TITLE OF INVENTION: Squalene Synthesis Enzymes
; FILE REFERENCE: BB1112 US CIP
; CURRENT APPLICATION NUMBER: US/09/909,745
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 60/107,241
; PRIOR FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 60/107,241
; PRIOR FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-909-745-24

Query Match 36.3%; Score 809.5; DB 10; Length 451;
Best Local Similarity 41.2%; Pred. No. 2.5e-67;
Matches 190; Conservative 76; Mismatches 146; Indels 49; Gaps 14;
QY 3 KAFSGKGAFLAGYIVLPTDYAVVTLSSRMVAHVPKQTSIKES---RIKISSPPQA 59
DB 5 RAFFSPGALLAGGLVLDYTRAEVGLSARMAVAHPYG-SLQGSKFEVRVSKPK 63
QY 60 NGMEWHYISNTB-KPEVQSRINPELFATFIYVLAIOPT-----EAPDLRIIYSD 111
DB 64 DGEMLYHISPRSGFIPVISIGSKNPFLEKIVANFYSYKPMMDYCNKNNLVID--IFSD 121
QY 112 PGYHSOEDTEKTSNGEKTFLYHSRAITEVEKTLGSSAGLVSVATSLSHFI---PN 168
DB 122 DAYHSOEDSVTE--HRGNRRILSPFHSRIIEVEPKTGLAGSLVTVLTTALASFVSDLEN 179

QY 200 TAIYGLIVRRPOPALINDV-----FOVLESDEKFPTELKLISSNEKKE 247
 DB 244 AAIYSGCLYRRFRSPILIESVDAGSPFERLERIYEDADPGHP-----MDTE-- 291
 QY 248 RC-----TTPYGLKLMGVDVKKSGSPFKYVSRVLOMKKEKPESSSVVYQOLANSALQFPMK 302
 DB 292 -CLDFGMKLPKGMQVLCVCEGSGQTPSMVRKYLEWRKQKQKADLWGALQSNMERLRL 350
 QY 303 ELREMERKXDSDEFTYIKELDSV---EPLTVAIKIRKGLQALQKSEVPIEPDVOQL 359
 DB 351 ELRRLAQSD-----EHTLSPFENVRITYIQSRNHISMVQKSDVPIEPVOTEL 400
 QY 360 LDRCEIPGCVGVVPGAGYDAIAYLVLENQ--VGNFKQ--KTLFN---PDY---FHHV 409
 DB 401 LIALSELGIVGVVPGAGYDAIYLLIQDNDVITRLKAFETMESKAEDEFGCGIGV 460
 QY 410 YVWDLSEQTEGVLEPEDEYIG 431
 DB 461 RLLGVHSGSEGVKNEMLEQYAG 482

RESULT 4

US-09-988-863A-2
 ; Sequence 2, Application US/0998863A
 ; Patent No. US20020123427A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bayer AG
 ; TITLE OF INVENTION: Plant phosphomevalonate kinases
 ; FILE REFERENCE: Ia A 35 018
 ; CURRENT APPLICATION NUMBER: US/09/988,863A
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 505
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 US-09-988-863A-2

Query Match 21.3%; Score 475; DB 10; Length 505;
 Best Local Similarity 28.3%; Pred. No. 7.9e-36;

Matches 143; Conservative 78; Mismatches 184; Indels 100; Gaps 15;

QY 1 MSKAFSPGKAFIAGGYLVLEPIYDAYVYATLSSRMHVAITPKGTSIKE-----SRIKI 53
 DB 1 MAVASAPGKVLMTGTYLVLEKPNAGLVLTNARFYAIKVPINEEKPESSAMKMTDVKL 60
 QY 54 SSPQFANGEMEYHISN--TEKPREVQSRINPFLBATTIFVLAYIQTEAPD----- 103
 DB 61 TSPQLSR-ESWYKLSLNHLTQSVASDSRNPFEHAIQYALAAHMLATEKOKESLHKL 119
 QY 104 ---LEIITYSDPGHSGED-----TEKTSNGEKTFLVHSR 137
 DB 120 LQGLDITTLGSDNFYSYRQIESAGLPLTPESLGTLPAPASITFNAASNG-----AN 172
 QY 138 AITVEKXTGLSSAGLVSVATSLSHFIPNYIS-----TNKDILHVAQIAH 185
 DB 173 SKPEVAKTGLSSAAMTTAVVAAL-HYI-GVVDLSDPCKEKGKFGCSDIDVHMTAQTS 230
 QY 186 CYAQKKGSGFVATAYIGLIVRRPOPALINDVEQVLESDE--EKFPTELKLISSNWE 243
 DB 221 CLAQKVGSGFVDSCAVYGSQRYVAFSPFVLSFAVVAATGLPLNEVITGLIK---GKMD 286
 QY 244 EKHRECTPIYGLKLMGDV-KGSETPKLYSVVLOMKKEKPESSSVYDQOLANSALQFPMK 302
 DB 287 NKRTFSILPPLMNLFLGEPGSGSSSTPBGVAVKQMSDPEKAKENNOQLSDANTLEET 346
 QY 303 ELREMERKXDSDEFTYIKELD-----HSEVPLTVAI-----KNIRKG 339
 DB 347 KLNDLSKLAKHMDVYLKIVKSCSVLTSEKVVLAHTEPINEALITELLAREAMLRITIL 406
 QY 340 LQALQKSEVPIEPDVOQLLDRCEIPGCVGVVPGAGYDAIAYLVLENQVGNFKQKT 399

DB 407 MEQGEAASVPIEPESQOTQLDSTMSASGVLAGVPGAGFDPAITLIGDSGKLTQAM 466
 QY 400 LBNPDYFHHVWVWDLSEQTEGVLEE 424
 DB 467 SS-----HNVALLVREDPHGVCLB 486

RESULT 5

US-10-036-959B-12
 ; Sequence 12, Application US/10036959B
 ; Publication No. US20030119098A1
 ; GENERAL INFORMATION:
 ; APPLICANT: E.I. du Pont de Nemours & Company
 ; APPLICANT: Hallahan, David L.
 ; APPLICANT: Keiper-Hyrynko, Natalie
 ; TITLE OF INVENTION: Genes involved in the Biosynthesis of Isopentenyl Diphosphate in
 ; FILE REFERENCE: CL-1792
 ; CURRENT APPLICATION NUMBER: US/10/036,959B
 ; PRIOR FILING DATE: 2002-05-10
 ; PRIOR APPLICATION NUMBER: 60/307,637
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 12
 ; LENGTH: 503
 ; TYPE: PRT
 ; ORGANISM: Hevea brasiliensis
 US-10-036-959B-12

Query Match 20.0%; Score 445.5; DB 15; Length 503;
 Best Local Similarity 27.4%; Pred. No. 4.6e-33;

Matches 138; Conservative 88; Mismatches 179; Indels 99; Gaps 16;

QY 1 MSKAFSPGKAFIAGGYLVLEPIYDAYVYATLSSRMHVAITPKGTSIKE-----SRIKI 53
 DB 1 MAVASAPGKVLMTGTYLVLEKPNAGLVLTNARFYAIKVPINEEKPESSAMKMTDVKL 60
 QY 54 SSPQFANGEMEYHISNTEKPREVQSRINPFLBATTIFVLAYIQTEAPD----- 103
 DB 61 TSPQLSR-ESWYKLSLNHLTQSVASDSRNPFEHAIQYALAAHMLATEKOKESLHKL 119
 QY 104 ---LEIITYSDPGHSGED-----TEKTSNGEKTFLVHSR 136
 DB 120 LQGLDITTLGSDNFYSYRQIESAGLPLTPESLGTLPAPASITFNAASNG-----AN 172
 QY 137 RAITEVEKXTGLSSAGLVSVATSLSHF-IPNYISTNK-----DILHVAQIAH 188
 DB 173 NCKPEVAKTGLSSAAMTTAVVAALHHLGLVDLSSECKEKKFSDLDVHIIAQTAHGA 232
 QY 189 QKITGSGFDVATAYIGLIVRRPOPALINDVFOVLESDEKFPTELKLISSNEKKE 248
 DB 223 QKVGSGFVDSAYVGSRRYVAFSPFVLSAQDAGKGP--LQVVISNLTGKMDHERBT 290
 QY 249 CTLPGYGLKLMGDV-KGSETPKLYSVRLOMKKEKPESSSVYDQOLANSAN-----LQFM 301
 DB 291 FSLPPLMSLILGEGTGGSSSTPBGVAVKQMSDPEKAKENNOQLSDANTLEET 350
 QY 302 KELEMERKXDS-----DPEYIKELDSVEPLTVAI-----KNIRKGLOA 342
 DB 351 SKLAE--EHMDAYKVIDSCSTKNSKKEIBQATEPSRAVVKALLGSNNMLQIHNRQ 408
 QY 343 LTQSEVPIEPDVOQLLDRCEIPGCVGVVPGAGYDAIAYLVLENQVGNFKQ--KTL 400
 DB 409 MGEAAGVPIEPESQOTRLDITMNDGVILAGVPGAGFDPAVATLIGDSGTNVAKWSL 468
 QY 401 ENPDYFHHVWVWDLSEQTEGVLEE 424
 DB 469 -----NVALLVREDPHGVCLB 486

RESULT 6

US-09-815-242-10656
 ; Sequence 10656, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FaastSeq for windows Version 4.0
 ; SEQ ID NO 10656
 ; LENGTH: 368
 ; TYPE: PRF
 ; ORGANISM: Enterococcus faecalis
 ; US-09-815-242-10656

Query Match 9.4%; Score 209.5; DB 9; Length 368;
 Best Local Similarity 27.0%; Pred. No. 4.1e-11;
 Matches 113; Conservative 52; Mismatches 137; Indels 117; Gaps 23;

6 SAVGKAFAGGYVLEPIYDAYVTALSSRMHVAITPKGSLKSRKISSPOPA-----59
 5 TTPGKLFPGYAVVBPGRHAIIVAD--QFVTVVEITDGS--IGSAQSSLPIM 59
 60 ---NGEMEVHISNTEKPREVQGRINPFLPATIPIYAVIOPTEAFDELIITVSDPGYS 116
 60 TRRNGELVLDI-----RPNPF-----HYVLAAILHLEKY-----A 89
 117 QETTERKTSNGKTYLHSHRAITVE-----KTGSSGSLVSVVATSLSHFIPNVIS 171
 90 QE-----QNKLSF-YHLKVTSELDSSNGRRTYGSS-GAVTVGVKALNIFYD--LG 138
 172 TKMDILHNAVQAHCQAOKKISGSPVATVY-GLIVRRFPQALINDVQVLESDEPKF 230
 139 LENEERFKLSALHNAVQGN-GSCGDIASCYGWAIFSTFDWDWN-----QKY 187
 231 PTE-LKTLIESNME-----KHERCTLPGIKILMGDVKGSETPKLVSRVLQMKKEP 283
 188 TRETLLDLAMDPELMIPFK-----VPKQALLLIGWTSPPASTSDIVDRVHQSKEK- 241
 284 EESSVYVDQNSANLQPMKELRMREKYSDEPTVYKELDHSVPEPLTVAIKRGLQAL 343
 242 -----QAAVEQFLMKSRLC-----VETMIN--GFNGKISVIOKQITKNROLL 282
 344 TOKSE--VPTEPDVQTOQLDRQOBIPGCVGVVPGAGVDAAVAVLENQVNFPROKT 399
 283 AEISLITGVITETALKNLCDLAESTYGAAS--SGAGGSDGIVL-----FRQKS 331

RESULT 7

US-09-909-745-22
 ; Sequence 22, Application US/09909745
 ; Patent No. US20020119546A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Falco, S. Carl
 ; APPLICANT: Famodu, Omolayo O.
 ; TITLE OF INVENTION: Squalene Synthesis Enzymes
 ; FILE REFERENCE: BBI112 US CIP
 ; CURRENT APPLICATION NUMBER: US/09/909,745
 ; CURRENT FILING DATE: 2001-07-20
 ; PRIOR APPLICATION NUMBER: 60/107,241
 ; PRIOR FILING DATE: 1998-11-05
 ; PRIOR APPLICATION NUMBER: 60/107,241
 ; PRIOR FILING DATE: 1998-11-05
 ; NUMBER OF SEQ ID NOS: 26
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 22
 ; LENGTH: 179
 ; TYPE: PRF
 ; ORGANISM: Oryza sativa
 ; US-09-909-745-22

Query Match 8.1%; Score 181.5; DB 10; Length 179;
 Best Local Similarity 31.5%; Pred. No. 5.5e-09;
 Matches 47; Conservative 21; Mismatches 56; Indels 25; Gaps 2;

264 GSETPKLVSRVLQMKKEPRESSVVDQNSANLQPMKELRMREKYSDEPTVYKELD 323
 1 GSSSTPSWGVSGVQKQSDPQSKETWSKLGANSVLNQLNKKLAEDHWEAYESVL- 59
 324 HSEVPLTV-----AIKNIRKGLQALTQKSEVPLEPVDQTL 359
 60 RSCSRLLTQSKTETVATNQHQLVNSLLARDAFLFELHMKREMGIAGVPIEPSQTL 119
 360 LDRCEIRPGCVGVVPGAGVDAAVAVL 388
 120 LDATNMEGVLAGVPGAGGFDVAFSVIL 148

RESULT 8
 US-09-815-242-5854
 ; Sequence 5854, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FaastSeq for windows Version 4.0

SEQ ID NO 5854
LENGTH: 345
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5854

Query Match
Best Local Similarity 7.8%; Score 173; DB 9; Length 345;
Matches 97; Conservative 57; Mismatches 143; Indels 128; Gaps 20;

7 AGKAFIAGGYLVLEPIYDAVYVATLSSRMHAYTP-----KGT-----SLKES 49
3 AGKGIYINGEYAVTEPGKSVLIALDRFVATIEETDQYKGIHKAHNPVTSRDE 62
50 RIKISSPOFANGEMWEHISNTEKREVOGRINPLEATIFVLAYIOPTAFDEIITY 109
63 SIVSDPHAAQ-----LNVVTAIEIFEGYAKSCDIAM--- 96
110 SDPGYHSQDETETKSSNGEKTFLYHSRAITEVEKGTGSSAG-LVSVAATSLSHPIPN 168
97 --KHFLITIDSNL-DDSNG-----HKYIGSSAAVAVSVI--KVINEFYDM 137
169 VISTNKDILHNVAOIAHCAOKKIGSGPDVATATY-GLIVRRPQPALINDVFOVLESBP 227
138 KLSNL--YIKLAIVANKMLQSLSSCG-DIAVSYSGLAVSTFHEWVKQIE----- 188
228 EKFPTELKLIENWEEKH-ERCTLPGYIKLMGDVKGSETPKLVSRVLOMKKEKPEES 286
189 ---DTTVEVLIKWPGLHIEPLQAPENMEVLIGWTGSPASPHVSEVKRLK----- 238
287 SVVTDQLSANLQFKEKLEKREKYSDEPTY---IKELDSVEPLTVAIK-NIRKGLQA 342
239 -----SDPSFYGDPLEDSHRCVEKLIHAFKNNIKGYOK 272
343 LTKSEVPIE-----PDVQTO---LIDRCQELPGCVGVPVPGAGYDAIAVLVLNQ 391
273 MVRQRTTIQRMDEKATYDIETEKLYCDIAEKYHG--ASKTSGAGGDC-GITTIKND 329
392 VGNFK 396
330 VDKER 334

RESULT 9

US-09-815-242-12866
Sequence 12866, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12866
LENGTH: 358
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12866

Query Match
Best Local Similarity 7.8%; Score 173; DB 9; Length 358;
Matches 97; Conservative 57; Mismatches 143; Indels 128; Gaps 20;

7 AGKAFIAGGYLVLEPIYDAVYVATLSSRMHAYTP-----KGT-----SLKES 49
6 AGKGIYINGEYAVTEPGKSVLIALDRFVATIEETDQYKGIHKAHNPVTSRDE 65
50 RIKISSPOFANGEMWEHISNTEKREVOGRINPLEATIFVLAYIOPTAFDEIITY 109
66 SIVSDPHAAQ-----LNVVTAIEIFEGYAKSCDIAM--- 99
110 SDPGYHSQDETETKSSNGEKTFLYHSRAITEVEKGTGSSAG-LVSVAATSLSHPIPN 168
100 --KHFLITIDSNL-DDSNG-----HKYIGSSAAVAVSVI--KVINEFYDM 140
169 VISTNKDILHNVAOIAHCAOKKIGSGPDVATATY-GLIVRRPQPALINDVFOVLESBP 227
141 KLSNL--YIKLAIVANKMLQSLSSCG-DIAVSYSGLAVSTFHEWVKQIE----- 191
228 EKFPTELKLIENWEEKH-ERCTLPGYIKLMGDVKGSETPKLVSRVLOMKKEKPEES 286
192 ---DTTVEVLIKWPGLHIEPLQAPENMEVLIGWTGSPASPHVSEVKRLK----- 241
287 SVVTDQLSANLQFKEKLEKREKYSDEPTY---IKELDSVEPLTVAIK-NIRKGLQA 342
242 -----SDPSFYGDPLEDSHRCVEKLIHAFKNNIKGYOK 275
343 LTKSEVPIE-----PDVQTO---LIDRCQELPGCVGVPVPGAGYDAIAVLVLNQ 391
276 MVRQRTTIQRMDEKATYDIETEKLYCDIAEKYHG--ASKTSGAGGDC-GITTIKND 332
392 VGNFK 396
333 VDKER 337

RESULT 10

US-09-909-745-23
Sequence 23, Application US/09909745
Patent No. US20020119546A1

GENERAL INFORMATION:
APPLICANT: Falco, S. Carl
APPLICANT: Famodu, Omolayo O.
TITLE OF INVENTION: Squalene Synthesis Enzymes
FILE REFERENCE: BBI112 US CIP
CURRENT APPLICATION NUMBER: US/09/909,745
CURRENT FILING DATE: 2001-07-20
PRIOR APPLICATION NUMBER: 60/107,241
PRIOR FILING DATE: 1998-11-05
PRIOR APPLICATION NUMBER: 60/107,241
PRIOR FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 26
SOFTWARE: Microsoft Office 97
SEQ ID NO 23
LENGTH: 378
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-909-745-23

Query Match
Best Local Similarity 5.5%; Score 123; DB 10; Length 378;
Matches 88; Conservative 74; Mismatches 145; Indels 130; Gaps 17;

```

QY 7 PGKAFIAGGYVLEBYIDYAYTALSSRMHNAVIT---PKGTSLKESRIKISSPOFANGEM 63
   ||||| |||||
Db 7 APGKIILIGEHAIV---VHGSTAVAAAIID.IYTVTLTRPLPSAENNDRLTL---QLOKDLSL 60
   |||||
QY 64 EYHISNTEKPREVOSRINPFLERATIFIVLAYIQPTEAFLEII.IYSDPHYSOE----- 118
   |||||
Db 61 BPSWS-----LARIKEAIPYDSSITLCTSPASGSEETLKSI 96
   |||||
QY 119 -----DTEKTSNGEKTPLY-HSRAI-----TIVE-KTGIGSSAGLVSV 157
   |||||
Db 97 AVLVEONLPEKERMLSSGISITFLMYLRIIGFNPAIVVINSHEL.PYSGSGSSAALCVAL 156
   |||||
QY 158 ATSLSHFI-----PNVISTNDIILHNVAQIACHYAOKI-----SGSPVATAIY 204
   |||||
Db 157 TRALLASSISEKTRGNMGWSSLDDETUELANKMA-----PEGEKIIHGKBSGIDINITYSAYG 211
   |||||
QY 205 LIVERYFOPALINDVFOVLESDEPEKFTPELKKLIESNWEKHERCTL.PYGIKLLMGDVKG 264
   |||||
Db 212 NMI-----KFCSEITRLQSN-----MP-LRMLINTTRV 239
   |||||
QY 265 GSEPEKIVSRVLQMKKEPESSVYVDQUNSAIQLQPMKLEKREKRYDSDPEYITIELDH 324
   |||||
Db 240 GNTATALVSGVQORAVRHPDANKSVBNADVISKELEALAIIOXSODERTSVTEKESRIEELME 299
   |||||
QY 325 SVEPELTVAIKNIRKGIQALOTKRSBVAIEBPDVQOTLDRCOEIPGCVGVPVGGAGYDAA 384
   |||||
Db 300 MNQGLLISNGVSHSISIEAV-----ILTTVHKHLVSK-----LTGAGGGGCVL 341
   |||||
QY 385 VLVLENQNGENFOKOTLE 401
   |||||
Db 342 TLL---PGTIVVDKVE 355
   |||||

```

```

RESULT 11
; Sequence 13670, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13670
; LENGTH: 335
; TYPE: PRY
; ORGANISM: Streptococcus pneumoniae
; US-09-815-242-13670

```

	Query March	5.5%	Score 122;	DB 9;	Length 335;	
	Best Similarity	22.8%;	Pred. No. 0.0059;			
	Match Local	97;	Conservative 52;	Mismatches 144;	Indels 132;	Gaps 21
Oy	9	GKAFLAGYLVLEPIYDAYVTALSSRMHAYITPKGTSIKES-RIKISSPQFANGEMEBYHI	67			
Dd	8	GKLWAGEYVAILLEGOLALINDIPYMRAEI-----AFSDSYRISDMDFDA-----	54			
Oy	68	SSNTEKPEVOSRNPN---FLEATFIYLAIY----OFTFAFDIEIIITYSDGYSQEDT	120			
Dd	55	-----YLRPPDYSLIQETIALMGDFLARGQLRPFSLEIC-----	92			
Oy	121	EKTTSNGEKTFLYHSRAITEVEKTKGSSAGLSVVATSLLSHEIPNVISTNKDILHNV	180			
Dd	93	-----GKMEREKKFGKGSSSVVLVVKALLATY--DVSDVGILLFGL	134			
Oy	181	AQIAHCYAOKTKIGSGFDVAATA-YGLIYYRRFPQRLINDVPQVESDEKPFTELKULE	239			
Dd	135	TS-VLLKRGDNGSNGLACTVAEDVLAYQSFR--QKVAAAMEE-----NATIVE	184			
Oy	240	SNNEKEHERCTLPYGAIKILMGDVKGSETPKLVSRYLQMKKEKPESSVVDOLNSANLQ	299			
Dd	185	RDW-----GFST--SQVKPLECDFVG--WTREVAVASHHVQQIQKNINQN	227			
Oy	300	FMKELREKREKYDDPEFYIK-----ELDHSEPLTYAINKIRKGLQALTQKSUVP	351			
Dd	228	FL-----SSKETTVSLVLEALECGKSEKIIEGVAVSK-LLEGIST-----	267			
Oy	352	EPDVOTLLDRCEQIEPGCVGGVV--PGAAGYDAIVLLENQVGNFKOKLTLENPDYFNHV	409			
Dd	268	--DIYTLPRLOLKESODLGAVASSGSAGGDCGIALSFDAQ---STKTLKN	314			
Oy	410	YWVDL	414			
Dd	315	RWADL	319			

```

RESULT 12
US-10-087-464-53
: Sequence 53, Application US/10087464
: Publication No. US20030059436A1
: GENERAL INFORMATION:
: APPLICANT: Chiehti, Alhar
: APPLICANT: Oh, Steven
: APPLICANT: Liu, David
: APPLICANT: Goel, Vikas
: APPLICANT: Li, Xuerong
: TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
: FILE REFERENCE: S1237/7019
: CURRENT APPLICATION NUMBER: US/10/087,464
: CURRENT FILING DATE: 2002-03-01
: PRIOR APPLICATION NUMBER: US 06/272,930
: PRIOR FILING DATE: 2001-03-02
: NUMBER OF SEQ ID NOS: 59
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 53
: LENGTH: 743
: TYPE: PRT
: ORGANISM: Plasmodium falciparum
US-10-087-464-53

Query Match      4.8%; Score 107.5; DB 15; Length 743;
Best Local Similarity 21.0%; Pred. No. 0.49;
Matches 82; Conservative 50; Mismatches 117; Indels 141; Gaps 17;

Oy      126  SNGKXTFLYHSRAITL-----VEKTLGSSAGLVSVATSL-----LSHFIPN 168
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      322  SNDLINFNLQSGKEIGENILNITLKNLGDKNNLSEFPDELIMLMDNLINYEFLDNLKTS 381

Oy      169  VISTNKDILHNVAOIAH-CYAAKKIGSGEDVATAIYGLIVRRFOPALINDVFOVLESDF 227
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      382  VLNKIKDILLRLLYKAAVYSYKKRAQE-----KGLP-----EPTVTN----- 418

```



```

Oy 177 LHNVAQAHCHYCAK-----KIGSGFVNAIAY-----GLIVRRF 211
Db 132 LKDAVYTAHNKIKENAIIGDEKEGTTLTFTAIYKQKCVIANCGDSRAVLLIDGSIIVERTK 191
Oy 212 QPALIN---DVFOULESDPEKPFTELKTLIES-----NWEKHRCUTLPJGIK 256
Db 192 DHSIVQVLVVEGHI SEEDARHPH -KKNITITALGIDPEKDDYEMD----- 236
Oy 257 LLMGDV--KQGSFETPKLVS--VLQWKKKKPRESSVWYDQULNSANLQFMKELRE 306
Db 237 LINGDVLNMSDGLHDIVSKEDILTKVYKNDHPKQIVBELFENTA-----LKTIRD 266

```

PRIOR FILING DATE: 2000-03-31
NUMBER OF SEQ ID NOS: 59
SOFTWARE: PatentIn version 3.0
SEQ ID NO 40
LENGTH: 1174
TYPE: PR
ORGANISM: Saccharomyces cerevisiae
US-10-841-40

Query Match 4.5%; Score 100; DB 15; Length 1174;
Best Local Similarity 22.0%; Pred. No. 5.2; 123; Indels 116; Gaps 21;

Matches 85; Conservative 63; Mismatches 123; Indels 116; Gaps 21;

QY 62 EWE-YHISNTEKPREVOSRINPF--LEATIFIVAIQPTFAFLEIIYSDPGYHSQ 117
DB 480 QMDLAKHI-----NEVDDIVPAEELKIDVLSNIFDTISAVEBELFFY--CFPL 529
QY 118 EDTETKTSNGEKTPLYHSRAITEVEKGTGSSAGLVSVATSLSHFIPNVIST----- 172
DB 530 REVISFDLDAKKKFFMRHFEILDEERKSNL--SYSVASKFATLVNNFTREQISLIDL 587
QY 173 -----NKDLH--NVAQIAHQAQ--KKISSGFDVATAIYGLI-----YRRF 211
DB 588 LANSTNLSLLKNDLIFERTNIT---YALINKLASVHQTFALALIQIPQICINKAV 642
QY 212 QPALIND-----VFVLESDEPEKFTTELKLTESNMEKHERCTLPYGIK 256
DB 643 RVALINNLTCESFCLODSATRECLHLSS-----PF-FKSNIFETN--YELCE-----K 688
QY 257 LMGDVKGGSETPKLVSRVLOWKKEKPESSVYVDOLANSANTQFMKELREMRKDSDE 316
DB 689 TMSPEMAISETGD-----EKKEIEDKISIFEKVMVTHLSQAKE--PVEKFELESY 738
QY 317 TYIKE---LDHSVEPLTVAIKNIRKGLQALTKSVPIPDVQYOLDRQCEIPGCVGV 373
DB 739 DIVKQSMISNGDSKLIINGFTIAKFL-----KPD-----NKRHDIQGMA--- 778
QY 374 VPGAGYDAIAVLVLNQNQNFQKTL 400
DB 779 -----ISYAKILENYSNFESERTI 798

RESULT 19
US-09-862-027-81
Sequence 81, Application US/09862027
Patent No. US20020142428A1

GENERAL INFORMATION:
APPLICANT: Hodge, Martin R.
TITLE OF INVENTION: No. US20020142428A1 Kinases and Uses Thereof
FILE REFERENCE: 35800/234862
CURRENT APPLICATION NUMBER: US/09/862,027
CURRENT FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: US 09/345,473
PRIOR FILING DATE: 1999-06-30
NUMBER OF SEQ ID NOS: 82
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 81
LENGTH: 1734
TYPE: PR
ORGANISM: Mus musculus
US-09-862-027-81

Query Match 4.5%; Score 99.5; DB 10; Length 1734;
Best Local Similarity 23.0%; Pred. No. 11; 125; Indels 143; Gaps 29;

Matches 99; Conservative 63; Mismatches 125; Indels 143; Gaps 29;

QY 28 VVALSRMAVITPKGT--SLKESRIKISSPOFANGEMEYHIS--NTEKREVOQRIN 82
DB 171 LALSPSGVGTPTPSSTVSSCSQEKHLQLPFOPTADLHLFLTGHFSTENVPDEGRRS 230
QY 83 PLEATIFIVLAIYQPTFAFLEII---YSD--PGYHSQDTE--TKTSSN----- 127
DB 231 PMRPRSRSLSPGRSV--SFDSIIIMMNVYKRFKATQAQMERPSLTFISSNTPDSVL 289

QY 128 ----GSEKELYHSRAITEVEKGTGSS--AGLVSVATSLSHRIPNVISTNKDILHNAQ 182
DB 290 PLADAGLSTFIHQ--VIEARCDLKSNSGLT-----SHVFTEL-----QENLEKIQ 336
QY 183 IAHCTAOKKIGSGFDVATAIY---GLIVRRFOPALINDVQVLESDEPEKTEPKLI 238
DB 337 DAHERSESS-----DVAFVIQVKKMLIAR--PARL-----LECLFDEEP--YHLEBA 385
QY 239 ESNMEKHERCTLPYGIKLMGDVKGSETPKLVSRVLOWKKEKPESSVYVDOLANSAL 298
DB 386 EGHAEKGH-----GIK-----CDIPRYIYQGLTLPDEMA-----QLSS--- 422
QY 299 QFMKELREMRKDYSDPEYIKELDHSVEPLTVAIKNIRKGLQALTKQSEVPLEPDVOT- 357
DB 423 -----YDS-PDI--PETDSVE-----GRVQSQPSQKT--PSEDFETI 456
QY 358 QLDRCQELIPGCVGVVPGAGYDAIAV-----LVLENQVN--FKQ 397
DB 457 KLISSN-----GAYGAVFLVRHKSTRQRFAMKKINKQNLIRNQIQQAFAVE 501
QY 398 KTL---ENP 403
DB 502 RDLTPAENP 511

RESULT 20
US-09-862-027-19

Sequence 19, Application US/09862027
Patent No. US20020142428A1
GENERAL INFORMATION:
APPLICANT: Hodge, Martin R.
TITLE OF INVENTION: No. US20020142428A1 Kinases and Uses Thereof
FILE REFERENCE: 35800/234862
CURRENT APPLICATION NUMBER: US/09/862,027
CURRENT FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: US 09/345,473
PRIOR FILING DATE: 1999-06-30
NUMBER OF SEQ ID NOS: 82
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 19
LENGTH: 394
TYPE: PR
ORGANISM: Homo sapiens
US-09-862-027-19

Query Match 4.4%; Score 99; DB 10; Length 394;
Best Local Similarity 20.1%; Pred. No. 1.1; 169; Indels 116; Gaps 17;

Matches 87; Conservative 60; Mismatches 169; Indels 116; Gaps 17;

QY 10 KAPLAGYLVL-----EPIYDAIVTALSSRMA-----VITPKGTSIKESRIKIS 54
DB 19 RAFWIGDEVAVAAARHDEDEDISQTIENVROEAKLPALMKHPIITALLRGVCLKEBNLCIV 78
QY 55 SPOFANGEMEYHISNTEKPREVQ-----SRINPFL--EATIPV-----LAYIQ 97
DB 79 M-EFARGPPLNRVLSGKRIIPDILVNAVQIARGMNVYHDAIVYIIHRDLKSSNILLIQ 137
QY 98 PTEAFDL--EIIYSDPGYHSQEDTETKTSNGEKTPLYHSRAITEVEKGTGSSAGLV 155
DB 138 KVENGDLSNKLITKLTDFGLAREMHRRTGMSAAG--TYAMMAPEVIRAFAMFSQDWSY 195
QY 156 VVATSLSHFIPNVISTNKDILHNAVAQIAHQAQKISSGFDVATAIYGLIVRRFOPAL 215
DB 196 VLLMELLTGEVP-----FRGIDGLRVAAGVAMNKLALPI- 229
QY 216 INDVQVLESDEPEKFTTELKLTESNME--EKHERCTLPYGI--KLMGDVKGSETPKLV 273
DB 220 -----PSTCFEPALMEDCNPDPHSPSFNIIIDLTTIESGFFEMPQDSF 278
QY 274 RVLQ--WKEKPESSVYVDOLANSANTQFMKELREMRKDYSDPEYIKELDHSVEPLTV 331
DB 279 HCLQDNWGHIEQ---MFDQLRAKE-----TWEEBELTRALQOKN 319

GENERAL INFORMATION:
APPLICANT: Boehringer Ingelheim Vetmedica GmbH
TITLE OF INVENTION: Recombinant attenuated PRRSV
FILE REFERENCE: 1-1120, 1125, 1126
CURRENT APPLICATION NUMBER: US/09/772,316
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: DE 10003371.7
PRIOR FILING DATE: 2000-01-26
PRIOR APPLICATION NUMBER: DE 10003373.3
PRIOR FILING DATE: 2000-01-26
PRIOR APPLICATION NUMBER: DE 10003372.5
PRIOR FILING DATE: 2000-01-26
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 2
LENGTH: 1457
TYPE: PRT
ORGANISM: Porcine reproductive and respiratory syndrome virus
US-09-772-316-2

Query Match 4.4%; Score 98; DB 9; Length 1457;
Best Local Similarity 21.2%; Pred. No. 11;
Matches 93; Conservative 65; Mismatches 178; Indels 102; Gaps 21;

QY 34 RHHAVTPKGTSLKESRIKISSPOFANGE-WEXHSSNTEKREPOGRINPLEATITIV 92
DB 952 RYGVVLTYPHRRREDDAITDSSOGAPFVVTLLPTKDSLNRGALVAITBARHAFVY 1011
QY 93 LAVIOTPEAFD-----EIIITYSD-----GYHSDTEKTSNGEKTFLYHSRAIT 140
DB 1012 DPHROQLGFLDPANGTPVNLAVHCDGLIVLDRNKECTVAALGNGKFRATDRVVD 1071
QY 141 EYEEK--TGLSSAGLVSVVATSLSHFIPN-----VISTNK-----DIL 177
DB 1072 SLRALCADLEGSSSPKPAHNLGFRSPDLQFAPLPELAPHMVSTONNEKPPDL 1131
QY 178 HNVAGIACYACQKIGSGFDVATAY---GLIVRRFQPALINDVFOYLESDEP----- 228
DB 1132 VASLPIHXYRACIGAGVWGPVFLTPGVVS---YLTKFVKGGAQVL-----PEVPEST 1186
QY 229 -KFFPELKLISNNEKREKCTLPYGIKLMGVKGS-----ETPLVSRVLTOMKKE 281
DB 1187 GRIVDCREYLD--REREVASLPRG---FIGDVGTGVGCHHTSRYLPRVL----- 1236
QY 282 KPESSVVDQINSANLQFMKELRERREKYDSDEPTYI-KELDHSVETPLVAIKNIR--- 337
DB 1237 -PKESVAVVGV--SSRGKAAKALCTITDYLPLEAYLHPETOSKCKMKMLDPKEVRLMV 1293
QY 338 -----KG-----LQALYORSEVPVIEPDVQTLDRCOELPGC-----VGVVPG 376
DB 1294 WHDKTAYFQLEGRVFTWYQLASVASYIRVFN---STVYLDPCWGPALCNRRVGSSTHMG 1350
QY 377 AG-----GYDAIVLVLEN 390
DB 1351 ADLAVTPYDYGAKITLSS 1368

RESULT 24
US-09-816-028A-17
Sequence 17, Application US/09816028A
Patent No. US20020042369A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Makarchuk, Warren W.
APPLICANT: National Research Council of Canada
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
FILE REFERENCE: 019633-000111US
CURRENT APPLICATION NUMBER: US/09/816,028A
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406

PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 17
LENGTH: 347
TYPE: PRT
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: beta-1,4 N-acetylglucosaminyl (GalNAc)
OTHER INFORMATION: transferase from C. jejuni strain CH4384 (ORF 5a)
OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus
US-09-816-028A-17

Query Match 4.3%; Score 96.5; DB 9; Length 347;
Best Local Similarity 18.2%; Pred. No. 1.6; Indels 117; Gaps 13;
Matches 65; Conservative 55; Mismatches 120;

QY 41 PKGTSLK-----ESRIKISSPOFANGEWEXHSSNTEKREPOGRINPF----- 84
DB 43 PKKTLVQINKNNEDLIKMLKALIGEGHKGYF--NYDEKSKOPKSPINMFARVNGEAI 100
QY 85 -LEATITIVLAIOTPEAFDEIITYSD-----PGHSDTEK 123
DB 101 TKKASLESILPAIQ-----RGVIGYNDCTDSEBIIIEFCQYPSFTPIKYPYEQION 154
QY 124 TSSNEKTEFLYHSRAITEVEKTGLSSAGLVSVVATSLSHFIPNVISTKDIHLNVAQI 183
DB 155 PKSEENKLSYV-----NYVASIFP-----KDEMLIKIDV 184
QY 184 AHCVAKKIGSGFDVATAYLGLIVRRFQPALINDVFOYLESDEPKFTELKLTESNWE 243
DB 185 DHIVDAKLYKSFYIPKKYDVYSRVDIHVFNNFLCKON-----NGNLT 232
QY 244 EKHECTLPYGIKLMGVKGSSETPKYSRV-LQMKKEKPESS----- 287
DB 233 KEGDCLLNNYNLMKREV-----LIDRINNWKATYQSFSSNHSLEOLKYKHRI 284
QY 288 VYDQINSANLQFMKELRERREKYDSDEPTYI-KELDHSVETPLVAIKNIRK 338
DB 285 LPHTELNNYHPPFLKKBRAODIYKYNWISIEFFKFFYONINHKLEPSMISKETLTK 341

RESULT 25
US-09-815-242-5644
Sequence 5644, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 1410
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5644
LENGTH: 496
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5644

Query Match 4.3%; Score 96.5; DB 9; Length 496;
Best Local Similarity 26.4%; Pred. No. 2.8;
Matches 67; Conservative 29; Mismatches 95; Indels 63; Gaps 14;

QY 53 ISSPQFANGEMWHISNTEKPREVOSRINPLEATFIYAVIOTPEADLEIIYS--110
DB 7 LSQRQYIDGEMWESANKNT---RDI---INPNQEVIFTV---SGTER-DAERALLAAR 56
QY 111 ---DPGYHSDTEKTSNGEKTFLYHRAITEVEKTLGSSAGLVSVVATSLSHFIP 167
DB 57 RAPESEMSQETAETR---GKIV---RAIADKIKERREALARLETLDTGKTLSESYA 107
QY 168 NVISTKDIHNVAAQIAHCYAOCKIGSGPD-----VATAI---YGLI- 206
DB 108 DM-----DIIHNVFMTFAGIADKDGEMIDSPIDTESKIYKEPVGVVQITPMYPLIQ 162
QY 207 VYRRQPALINDVFOVLESDEPEKFTPELTKLLISNWEKHERCTLPYG-IKILM----- 259
DB 163 ASWKIAPALATGCSLVK-----PSEITPLTIRVFELMEBVGFPKGTINILGAGSEV 216
QY 260 GDVKGSETPKLYS 273
DB 217 GDVMSGHKEVDLYS 230

RESULT 26
US-09-815-242-12657
Sequence 12657, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12657
LENGTH: 496
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12657

Query Match 4.3%; Score 96.5; DB 9; Length 496;
Best Local Similarity 26.4%; Pred. No. 2.8;
Matches 67; Conservative 29; Mismatches 95; Indels 63; Gaps 14;

QY 53 ISSPQFANGEMWHISNTEKPREVOSRINPLEATFIYAVIOTPEADLEIIYS--110
DB 7 LSQRQYIDGEMWESANKNT---RDI---INPNQEVIFTV---SGTER-DAERALLAAR 56
QY 111 ---DPGYHSDTEKTSNGEKTFLYHRAITEVEKTLGSSAGLVSVVATSLSHFIP 167
DB 57 RAPESEMSQETAETR---GKIV---RAIADKIKERREALARLETLDTGKTLSESYA 107
QY 168 NVISTKDIHNVAAQIAHCYAOCKIGSGPD-----VATAI---YGLI- 206
DB 108 DM-----DIIHNVFMTFAGIADKDGEMIDSPIDTESKIYKEPVGVVQITPMYPLIQ 162
QY 207 VYRRQPALINDVFOVLESDEPEKFTPELTKLLISNWEKHERCTLPYG-IKILM----- 259
DB 163 ASWKIAPALATGCSLVK-----PSEITPLTIRVFELMEBVGFPKGTINILGAGSEV 216
QY 260 GDVKGSETPKLYS 273
DB 217 GDVMSGHKEVDLYS 230

RESULT 27
US-10-171-311-2
Sequence 2, Application US/10171311
Publication No. US20030087270A1
GENERAL INFORMATION:

APPLICANT: Schlegel, Robert
APPLICANT: Chen, Yan
APPLICANT: Zhao, Xumei
APPLICANT: Monahan, John
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Glatt, Karen
APPLICANT: Gannavarapu, Manjula
APPLICANT: Hoersch, Sebastian
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
FILE REFERENCE: MRI-035
CURRENT APPLICATION NUMBER: US/10/171,311
CURRENT FILING DATE: 2002-06-12
PRIOR APPLICATION NUMBER: US 60/298,159
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,155
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/335,936
PRIOR FILING DATE: 2001-11-14
NUMBER OF SEQ ID NOS: 238
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 3907
TYPE: PRT
ORGANISM: Homo sapiens
US-10-171-311-2

Query Match 4.3%; Score 96; DB 15; Length 3907;
Best Local Similarity 20.4%; Pred. No. 86;
Matches 88; Conservative 69; Mismatches 180; Indels 94; Gaps 17;

QY 35 MNAVITPKGTSLKESRIKISSPQFA-----NGMEWHISNTEKPREVOSRINPLEA 87
DB 1810 MMSKTEBETELSOBLVSG---FAGTEIDPENEBELMUNISS-----RLQAAVEKLEA 1860
QY 88 TFIYAVIOTPEADLEIIYSDEPGYHSDTEKTSNGEKTFLYHRAITEVEKTL 147
DB 1861 -ISETSSQLEHAKVQTGLM--RESFROKQKATESLKCOEERLHESRRARQOLAVEL 1917
QY 148 GSSAGLVSVVATSLSHFIPNVISTKDIHNVAAQIAHCY-----QKIGSGPDV 198

Db 1918 SBAEGVIDYADE--KTLEFROIQEKTDIIDLBOELLCASNRLOELAEQOIOEEREL 1975
Qy 199 ATAIVGLIVRRFOPALINDVFOVLESDEPEKTELTCLIESNWEKHERCTLPYGIKL 258
Db 1976 -----LSRQKAMRAEAPVQOQLLOETEKL---MKELEVOQCAEKVRDLOKOVKL 2026
Qy 259 MGDVKGSETPLVSRVLOMKKEKREPRESSVYVDQNSANLQPMKELREMERKYDSDP--- 315
Db 2027 EIDVE-----EQVSRFIELEQKNTLEL---DLRQOQALEKQLEKRRKFLDEQALDR 2076
Qy 316 -----ETVKELDHSV-----EPLTVAIKIRKIGLOALTK--SEVPIDP-VQ 356
Db 2077 EHERDVFQOEIQLEQOLKVVRFPQIPSEHQTRREVQOLANHLKERTDKCSSELLLSKEQOL 2136
Qy 357 TOLLRCOEIPGCVGVVPGAGYDAI AVLVLNVOGNFKQKTLNPDYFNHVVWDLBE 416
Db 2137 RDIQERNEIE-----KLEFVRLEQALIVSADTFQVBE-----DR 2173
Qy 417 QTEGVLEEKPE 427
Db 2174 KHFGAVEAKPE 2184
RESULT 28
US-10-171-311-6
; Sequence 6, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatz, Karen
; APPLICANT: Gamaavarapu, Manjula
; APPLICANT: Hoerzh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3925
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-6
Query Match 4.3%; Score 96; DB 15; Length 3925;
Best Local Similarity 20.4%; Pred. No. 86; Indels 180; Gaps 17;
Matches 88; Conservative 69; Mismatches 180; Indels 94; Gaps 17;
Qy 35 MAAVTPKGTSLKESRIKISSPOFA-----NGEMEYHISSTNTEKPREVQSRINPFEA 87
Db 1810 MMSKVTBETSTELSGQLVSG---FAGTEIDPENBELMINISS-----RLQAAVEKLEA 1860
Qy 88 TTFIYLAIOPTPEADLEIIYSDPGYHSQEDTEFTKTSNGEKTLYHSRAITEVEKTL 147
Db 1861 -ISSTSSQLEHAKVTQTELM--RESFRQKQZATSESLKCEBELRLHESRAREQLAVEL 1917
Qy 148 GSSAGLVSVATSLSHFIPNVISTNKDILHVAQIAHCA-----OKKISGPDV 198
Db 1918 SBAEGVIDYADE--KTLEFROIQEKTDIIDLBOELLCASNRLOELAEQOIOEEREL 1975
Qy 199 ATAIVGLIVRRFOPALINDVFOVLESDEPEKTELTCLIESNWEKHERCTLPYGIKL 258

Db 1976 -----LSRQKAMRAEAPVQOQLLOETEKL---MKELEVOQCAEKVRDLOKOVKL 2026
Qy 259 MGDVKGSETPLVSRVLOMKKEKREPRESSVYVDQNSANLQPMKELREMERKYDSDP--- 315
Db 2027 EIDVE-----EQVSRFIELEQKNTLEL---DLRQOQALEKQLEKRRKFLDEQALDR 2076
Qy 316 -----ETVKELDHSV-----EPLTVAIKIRKIGLOALTK--SEVPIDP-VQ 356
Db 2077 EHERDVFQOEIQLEQOLKVVRFPQIPSEHQTRREVQOLANHLKERTDKCSSELLLSKEQOL 2136
Qy 357 TOLLRCOEIPGCVGVVPGAGYDAI AVLVLNVOGNFKQKTLNPDYFNHVVWDLBE 416
Db 2137 RDIQERNEIE-----KLEFVRLEQALIVSADTFQVBE-----DR 2173
Qy 417 QTEGVLEEKPE 427
Db 2174 KHFGAVEAKPE 2184
RESULT 29
US-09-815-242-10779
; Sequence 10779, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyckind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10779
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10779
Query Match 4.3%; Score 95.5; DB 9; Length 930;
Best Local Similarity 22.1%; Pred. No. 9.4; Indels 77; Gaps 19;
Matches 85; Conservative 58; Mismatches 165; Indels 77; Gaps 19;
Qy 30 AASRMNAV---TTPKGTSLKESRIKISSPOFAN--GEWEYHISSTNTEKPREVQ 79
Db 169 SINEKINAIAPKITSKASISQOISEBFIKTASTLTKTENDIGYDIDKMWVSIOKYKS 228
Qy 80 RINPLEKTIYVLAAYIOPTPEADLEIIYSDPGYHSQ--BDTEKTSNGEKTFLYHSR 137
Db 229 MILD-TDNNICTITPYA--OVTDL-----HGMPLEKEDLAVANAMKTLPEVD 275
Qy 138 AITE--VEKTEG-LSSAGLVSVATSLSHFIPNVISTNKDILHVAQIAHCAQAKIGS 194

Db 276 ALGKELVINGKPSIKEQASVILT--LOEKIPEIONAGROI-----AMIDEDPASYEQTM 329
Qy 195 GPDVAIAIYGLIYRRFPQAL-----INDVFOVLESDPEKFTTELKLIESWEEK 245
Db 330 SEGIDAKOGLETIIQOVQTLPIRIKLGQANDLGAVTLGANKLOEALPSITNS----- 384
Qy 246 HERCTLPYGIKILMGVKG-GSETPKLVSRVLQ-----WKKEPSESSVVD----- 291
Db 385 -----VEVLKSIQVATTTTSVATIRQALDDGULTPEEKOHINEVODFTTNIQ 435
Qy 292 -QINSAN--LOFMKELREMEKXYDSDETYIKELHDSVEPLVAIINIRKGLQALTQKSE 348
Db 436 ROQQAINDITAFMKOLQENAGNHDLDGA--IAQLSH-VNDLLTDFSNRLNQLNALVQAGD 492
Qy 349 VPIEPVQVQTLDRCEIPECVGCV 373
Db 493 ISGVQNYLNEIDEMATNISSIGSV 517

RESULT 30

US-10-171-311-4
; Sequence 4, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamakkar, Shubhangi
; APPLICANT: Glatte, Karen
; APPLICANT: Ganavathapu, Manjula
; APPLICANT: Hoerth, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 3899
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-4

Query Match 4.3%; Score 95.5; DB 15; Length 3899;
Best Local Similarity 21.3%; Pred. No. 95;
Matches 94; Conservative 68; Mismatches 163; Indels 117; Gaps 20;
Qy 39 ITPKGTSLKSRKIKISSPOFANGEMEHYISNTEKREVOGRINPF---LEATIFIVLAY 95
Db 3118 IQQKQSQMELQVLEISMKDRATELOQLSSKRVAAELKSLAQTKLELTLLKQKHG 3177
Qy 96 IOPTEAFDELEIITYSDPGHSDPTETKTSNGEKTFLYHSRAI---TEVEKTGLGSSAG 152
Db 3178 LKLELAFRLVEXDKTDE-VHLANDT---LASBQK-----SRELQWALEKEKAKLGRSE- 3227
Qy 153 LVSVAATSLSHFIPNVISTNKDILHNAQIAHCVAAQKIGSGPDVATAIYGLIYVRRQ 212
Db 3228 -----ERDKLEELDLKFLSLSQKQKRNVL-----QNLILLEQ 3258
Qy 213 PALINDVFOVLESDPEKFT-----ELKKLIESWEEKH-RCTLPYGIKL---L 258
Db 3259 KOLINESQOKIESQRMLYDAQISEEGRNLEQVLESEKVRIRMSSTLDRERELHAOL 3318

Qy 259 MGDVKGSETPKLVSRVL--QMKKEPSESSVVDQINSANLQFMKELREMEKXYDSDE 316
Db 3319 QSSDGTGQSRPPLPSEDILKELQKOLEEKHSRIVELNTE-KYKLDLSIQTRQOMEKDRQ 3377
Qy 317 TYIKELHDSVEPLVAIINIRKGLQALTQKSEVPIEPVQVQTLDRCEIPECVGCVAG 376
Db 3378 VHRKTLQTRQENANTG---QKRMHLOSRYE---DLQRLERKROQ----- 3417
Qy 377 AGGYDAIAVLVLNQ-----VGNFKQTLNPD-----YFHV----- 410
Db 3418 -----VYKLDLEGGRLQSGIMQEFQKOLEERKEKESRILYQNLNEPTTMSLTSDRTRN 3471
Qy 411 WVDLEQTEGVLEEKPEDEYGL 432
Db 3472 WV-LOQKIEG--ETKESNYAKL 3490

RESULT 31

US-10-171-311-8
; Sequence 8, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamakkar, Shubhangi
; APPLICANT: Glatte, Karen
; APPLICANT: Ganavathapu, Manjula
; APPLICANT: Hoerth, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 3917
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-8

Query Match 4.3%; Score 95.5; DB 15; Length 3917;
Best Local Similarity 21.3%; Pred. No. 96;
Matches 94; Conservative 68; Mismatches 163; Indels 117; Gaps 20;
Qy 39 ITPKGTSLKSRKIKISSPOFANGEMEHYISNTEKREVOGRINPF---LEATIFIVLAY 95
Db 3118 IQQKQSQMELQVLEISMKDRATELOQLSSKRVAAELKSLAQTKLELTLLKQKHG 3177
Qy 96 IOPTEAFDELEIITYSDPGHSDPTETKTSNGEKTFLYHSRAI---TEVEKTGLGSSAG 152
Db 3178 LKLELAFRLVEXDKTDE-VHLANDT---LASBQK-----SRELQWALEKEKAKLGRSE- 3227
Qy 153 LVSVAATSLSHFIPNVISTNKDILHNAQIAHCVAAQKIGSGPDVATAIYGLIYVRRQ 212
Db 3228 -----ERDKLEELDLKFLSLSQKQKRNVL-----QNLILLEQ 3258
Qy 213 PALINDVFOVLESDPEKFT-----ELKKLIESWEEKH-RCTLPYGIKL---L 258
Db 3259 KOLINESQOKIESQRMLYDAQISEEGRNLEQVLESEKVRIRMSSTLDRERELHAOL 3318
Qy 259 MGDVKGSETPKLVSRVL--QMKKEPSESSVVDQINSANLQFMKELREMEKXYDSDE 316
Db 3319 QSSDGTGQSRPPLPSEDILKELQKOLEEKHSRIVELNTE-KYKLDLSIQTRQOMEKDRQ 3377

TYPE: PRT
ORGANISM: Hevea brasiliensis
US-10-036-959B-11

Query Match 4.3%; Score 95; DB 15; Length 386;
Best Local Similarity 20.0%; Pred. No. 2.6;

Matches 83; Conservative 72; Mismatches 151; Indels 108; Gaps 19;

QY 7 APGKAFIAGYLVLEPIYDAVYVTLASRMMAVITPK-GTSLKESRIKISSPOFANGSEW 65
DB 7 APEKIIISGEHAY---VHGSTVAASAINITYTTLSPATLENDSLKQLKDAL-ERFW 62
QY 66 HHS-----SNTKEPREVOGRINPLEATTIYVLAIVIQTEAPDELIIYSDGYSQED 119
DB 63 PIGIRREALSNLAPSS-STRTSCSME-SIKTISALVEENIPEAKIAL----- 109
QY 120 TETKTSNGEKTLYHSRATVEK-----TGSSNAGLVSVATLSLHFTP 167
DB 110 -----TSGVSAPFLMYTSTIQGFKPATVVTSDLPLOSGSGSSAAFCVALSALLAF--- 160
QY 168 NVISTNKDILH-----NVAQIAHCYA---OKKI---GSGPDVATATVGLVYRFP 213
DB 161 -SDSVAVDTHTLMSIFGESDDELKMKALEGKIIHGKPSGIDNTVSAYGMMI----- 213
QY 214 ALINDVFOVLESDEPEKPEPELKLISNMEKEHERCTLPYGIKILMGDVKGSETEPLVS 273
DB 214 -----KFKSGNLTRIKSN-----MP--LKMIVNTRVGKNTALVA 247
QY 274 RVLIQMKKEKEPSESVYVDOLNSANLQPMKELREKRYDSDEPTTYIKELDHSVEPLTVAI 333
DB 248 GVSERTLRHFNMAFSFENAVDSIS---MELANIIOSPAPD-DVSTKEKEKLEEL----- 298
QY 334 KNIRKIGQALTOQSEVPEPDVOTOLLDRQCQIPGCGVVPAGVYDAIVLV 387
DB 299 MEMNGHILQCMGVSHASIEVTYRTLYKX-----LASKULTAGAGCCVLTLL 345

RESULT 35

US-10-156-761-10884

Sequence 10884, Application US/10156761

Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 10884
LENGTH: 478
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-10884

Query Match 4.3%; Score 95; DB 15; Length 478;
Best Local Similarity 20.6%; Pred. No. 3.6;

Matches 59; Conservative 47; Mismatches 98; Indels 82; Gaps 11;

QY 197 DVATATYGLIVRFPALINDVFOVLESDEPEKPEPELKLISNMEK-----HR 248
DB 18 DIAPS-YGLFIDSEFLBADGVFKVTSSTSEVLSEIAQAGADVRAVKARKAPEK 76
QY 249 CTLPGIKILMGDVKGSETEPLVSRVLQWKKEKEPSESVYVDOLNSANLQFMKELREMR 308

DB 77 SALP-----GSERAKYLFRIARIQERREILAV-----LETLDNKRPXK 115
QY 309 EKYDSDEP-----YIKELDH-----SVEPLTVAKNIR-----KGLQAL 343
DB 116 ETRDADLPVLAHFFYYAGMDKLDAGGANPRPLGVAGQVYPMNPLMLAMKIAPL 175
QY 344 TOKSEVPIEPDVOTOL-----LDRQCE--IPGCVGVVPGAG----- 378
DB 176 ATGNVTYVLKPAFTTPLSALFFADICRQAGLPKGVNIIIPGYDAGALVEHPDVKNVAF 235
QY 379 GYDAIAVLLENQVNFKOKTLENPDYFHNVTWVD--LEEQTEGVL 422
DB 236 GSTAVGAKAIARQVAGTEKKTLELGKGANIVFDAPIDQAVEGIV 281

RESULT 36

US-09-925-300-1219

Sequence 1219, Application US/09925300

Patent No. US20020151681A1

GENERAL INFORMATION:

APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben,
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925,300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1219
LENGTH: 590
TYPE: PRT
ORGANISM: Homo sapiens

FEATURE:
NAME/KEY: SITE
LOCATION: (17)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (116)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (127)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (131)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (134)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (158)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (161)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (213)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (216)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1219

Query Match 4.2%; Score 94.5; DB 10; Length 590;
Best Local Similarity 20.7%; Pred. No. 5.6;

Matches 75; Conservative 39; Mismatches 117; Indels 131; Gaps 16;

QY 135 HSRATTEVEKTGLSGAGLVSVATSLSHPIPNVISTNKDILHNAQIAHCYAKKIGS 194

Db 160 HKRSAREV-----IAVMD-----VFTDIDIRDOEI--CRQ----- 191
 Qy 195 GPDVATAIYGLIYRRFOPALIN--DVEQVLESDPEKEPTELEKLIEN-- 241
 Db 192 -----GVANYILLDQALLSQFLDMCKMLKQHEQEKMTVRRITIGNIYARSGTKI 242
 Qy 242 WEEKHERCTLPYGIKILMG-----DVKGSE----- 267
 Db 243 IGVHHEKFTLLIDIRVATGSSYSFTWTDGKLSNVLISGQVVEHFLEFRILYAQSKPI 302
 Qy 268 TPKLVSRLVQWKK-----EKPESSVYVDQLSANLQPMK--ELAREMEKXDSDET-- 317
 Db 303 SPKLSSHFGSSNKFHDHITNRKPSK-----ELTLGNILRRRLARLSSTPRKADLDEMPA 357
 Qy 318 --YIKELDSVEPLTYAIKNI-----RKGLQALTK--SEVP-----IEPDVOTQ 358
 Db 358 EGAKEKRPDCSESTYSEEDYSESHDELQSKRAIQAOTBEGEEMPGLSSEVGTQIS 417
 Qy 359 LIDRCOEIPGCVGVVAGGYDAIAVLVLEQVGNFKQTLLENPDYFHNVYVWDLEROT 418
 Db 418 ITTAC-----AGTCAVITRIASSQTITMSSTTTQTDMDENILFPRTGOST 464
 Qy 419 EG 420
 Db 465 EG 466

RESULT 37

US-09-828-062-8
 ; Sequence 8, Application US/09828062
 ; Publication No. US20030097675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: COSTA E SILVA, OSWALDO DA
 ; APPLICANT: BOHNET, HANS J.
 ; APPLICANT: VAN THIELEN, NOCHA
 ; APPLICANT: CHEN, RUTYING
 ; APPLICANT: SARRIA-MILLAN, RODRIGO
 ; TITLE OF INVENTION: CELL CYCLE STRESS-RELATED PROTEINS AND METHODS OF USE
 ; FILE REFERENCE: 16313-0031
 ; CURRENT APPLICATION NUMBER: US/09/828,062
 ; CURRENT FILING DATE: 2001-08-20
 ; PRIOR APPLICATION NUMBER: 60/196,001
 ; PRIOR FILING DATE: 2000-04-07
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 8
 ; LENGTH: 901
 ; TYPE: PRT
 ; ORGANISM: Physcomitrella patens
 US-09-828-062-8

Query Match 4.2%; Score 94; DB 11; Length 901;

Best Local Similarity 18.2%; Pred. No. 12;

Matches 74; Conservative 55; Mismatches 100; Indels 178; Gaps 20;

Qy 2 SKAFSAPGKAPLAG--GYLVLEPIYDAYVTALSRHNAVITPKGSLKESRIKISSPOA 59
 Db 78 NEVYSSPYAGTPGTPGTATPVY-----ATPVGTPM-----GTPSPH 116
 Qy 60 NGEWEYHSS--NTEKP--REVQSRINPLEATIFIVLAVIOPTENADELIIYSDP 112
 Db 117 RCTPGYKQSELSGQCKPLHRRRSQSR-----GTVNV-----EP 146
 Qy 113 GYHS-----QEDTETSSNGEKTFLYHSRAITEVEKTLGSSAGLVSVANS 161
 Db 147 GHRSPRESADGRPSESAPDDTLGGEYAYW-----GTVNV----- 184
 Qy 162 LSHFIPIVISTNKDILHNVAQIAHCYAOKKISSGPDVATAIYGLI--VYRRFOPALINDY 219
 Db 165 ----IPDVRAIRRLPLANTRSSAH-----DANSKYIQLIEVYEREDTLNIDM 229
 Qy 220 FQVLESDEP-----KFTLEKLKLIENME-----EKHERCTLPYGIKL----- 257

Db 230 SDIYHDPDLVAKIYVYRPIIDILDLTDECQEVATSLPTEFKIIEAR--FENLKASVYMR 288
 Qy 258 -----LMGVKGS-----SEPKVLSRVL-----QMKEXE 284
 Db 289 LNPSSIDIKLVSKGVIRCSSIIPETKAPFKCLVCGSHPPPLVTVKGVVEEPTCEK 348
 Qy 285 -----ESSVYVDQLSANLQPMKELREMEKXDSDETEYIKELDSHV 326
 Db 349 CAARNMSLIHNRCTPAN-----KQIVRLQETDAIPE--GTPHTV 388

RESULT 38

US-09-820-843A-16
 ; Sequence 16, Application US/09820843A
 ; Publication No. US20030039963A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Council of Scientific and Industrial Research
 ; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
 ; FILE REFERENCE: 063915
 ; CURRENT APPLICATION NUMBER: US/09/820,843A
 ; CURRENT FILING DATE: 2001-03-30
 ; NUMBER OF SEQ ID NOS: 118
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 16
 ; LENGTH: 1616
 ; TYPE: PRT
 ; ORGANISM: M. genitalium
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: cyc adherence-accessory protein
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: gi1046097
 US-09-820-843A-16

Query Match 4.2%; Score 94; DB 11; Length 1616;

Best Local Similarity 21.9%; Pred. No. 32;

Matches 113; Conservative 63; Mismatches 161; Indels 180; Gaps 29;

Qy 48 ESRITSSPPQFANGEM--EYHISNTEKPREVQSRINPLEATIFIVLAVIOPTENAFD-- 103
 Db 605 ETEVAVSS--ELPKSLVDEITFINNDPKQEGLEKVDPLET-----EPKSLFDEK 654
 Qy 104 LEIITYSDPGY-----HSQEDTETKTS--NGK--TFL--YHSRAITE 141
 Db 655 TTIVSESPPTIQDPLSLSDSVNDVDKSLKTKTSVBLNHEIGNEFTINDVSEKVOE 714
 Qy 142 VEKTLGSSAGLV--SVVATSLT--SHFIPNVIST--NKDILHNVAQIAHCYAOKKISSG 195
 Db 715 OPTQLDTSERFVLPYQIVEDSFTESSAETREPFSSQKDTLEFISQ-----TOEVETSE 769
 Qy 196 PDVATAIYGLIYRRFOPALINDVFOVLESDEKEPTELEKLIEN--WEEKHERCTLPY 253
 Db 770 SNVPVEQET--KLFEHQDENNLFTPL-----PLDLTEIIESNALFDSK----- 811
 Qy 254 GIKLIMGVKGSSER--PKVLSRVLQMKKEKPESS-----VYVDQLSANL 298
 Db 812 -----PDEKSSDSELOPTFEKIKLDSVTEVPQSSQVEATETVQPEAVVEDEIKTOEL 865
 Qy 299 Q-----FMKEIREMREKRYDSDPE--TYIKELDSVEPLTYAIKNIK-- 338
 Db 866 QPEATTEVPDHPQPDVQPEGTPEAKFDSVEIIPQSSQAFPAEQISDIKLEKTE 925
 Qy 339 ---GLQALTKSEVPI-----EPD--VOTQ-----LDRQOEIPGCVGV 373
 Db 926 AVFDHQLEQNSSEETVVPTEVTAPEPETIETOLEPSSSEDQPSBPALD--QNHPEIYAE 983
 Qy 374 VPGAGYAIIVLLEQVGNF-----KOKT----- 399
 Db 984 VEQI--FDGTLEDIKLEANFNDVNEVOPKETEAITETDEKELQOETSSEPLSTEE 1041
 Qy 400 LENPDYFHNVYVWD-----LEEQTEGVLEEKPE 427

Db 1042 LKSEATFDNVSEASEAVEFEKPEQLETKLEEEPK 1078

RESULT 39

US-09-987-107-33

/ Sequence 33, Application US/09987107

/ Patent No. US20020156007A1

/ GENERAL INFORMATION:

/ APPLICANT: GRAVERSEN, Jonas

/ APPLICANT: MOESTRUP, Soren

/ TITLE OF INVENTION: APOLOPROTEIN ANALOGUES

/ FILE REFERENCE: GRAVERSENIA

/ CURRENT APPLICATION NUMBER: US/09/987,107

/ PRIOR FILING DATE: 2001-11-13

/ PRIOR APPLICATION NUMBER: US 60/264,022

/ PRIOR FILING DATE: 2001-01-26

/ PRIOR APPLICATION NUMBER: DK PA2001 00057

/ PRIOR FILING DATE: 2001-01-15

/ PRIOR APPLICATION NUMBER: DK PA2000 01682

/ PRIOR FILING DATE: 2000-11-10

/ NUMBER OF SEQ ID NOS: 91

/ SOFTWARE: Patent version 3.1

/ SEQ ID NO 33

/ LENGTH: 396

/ TYPE: PRT

/ ORGANISM: Homo sapiens

US-09-987-107-33

Query Match 4.2%; Score 93.5; DB 10; Length 396;

Best Local Similarity 20.1%; Pred. No. 3.7;

Matches 80; Conservative 63; Mismatches 155; Indels 101; Gaps 18;

QY 38 VITPKGSLKESRIKISSPOFANGWEY--HISNTE-----KPEVQSRINPLEATI 89

Db 7 VLTALVAVAGARAEVSADQVATVMWDYFSQLSNNKAEVHLQKSELTOQLNALFQDKL 66

QY 90 FIVLAIQPTAEADLE--IIYSDPGYHSQEDTETKTSNGEKTFLYHSRAITEVEKTGL 147

Db 67 GEVNTY-----AGDLQKLVPPA-----TE-----LHERLADSEK--L 98

QY 148 GSSAGL-VSVVATSLSHFIPNVISTN-KDILHNAQIAHCYA-QKKIGSGFDVATAIYG 204

Db 99 KEIGKELEELRRLPH--ANEVSQKIGDNLRELQORLEPYADQLRQVNTQAEQLRQ 156

QY 205 LIYRRFPQPALINDVQVLESDEPEKPEPTELKLIENMEKHEKRCPLPYGIKLMGDVYG 264

Db 157 LTPYAQRMERVLRENADSLQASLRPHADELAKKIDQVVELKGRLT-PYA----- 205

QY 265 GSETPKLVSRLQWKKEKEPESSVYVDQLSANLQFMKELREMERREKYSDPEPTYIKELDH 324

Db 206 -----DEFKVKIDQT-----VEELRSLAPYADQ-----EKLNH 236

QY 325 SVEPLTVAIKNIRKGLQALTQKSEVPPIEPDVOTQLDRQCEIPGCVGVPG-----A 377

Db 237 OLEGLTFQMKNAEELKARISAS---AEELRQRLAPLAEDVGNLKGNTGELQKSLAEL 292

QY 378 GGYDAIAVLVLNVOVGNFKQKTLNPDYFHNYYWVLEE 416

Db 293 GGH-----LDQVBEFFRRRVEPYGENFKALVQOMEQ 324

RESULT 40

US-09-802-640-16

/ Sequence 16, Application US/09802640

/ Publication No. US20030036057A1

/ GENERAL INFORMATION:

/ APPLICANT: Braun, Andreas

/ APPLICANT: Kiegn, Patrick

/ APPLICANT: Kiegn, Patrick

/ TITLE OF INVENTION: GENES AND POLYMORPHISMS ASSOCIATED WITH

/ TITLE OF INVENTION: CARDIOVASCULAR DISEASE AND THEIR USE

/ FILE REFERENCE: 24736-2048

/ CURRENT APPLICATION NUMBER: US/09/802,640

/ CURRENT FILING DATE: 2001-03-09

/ NUMBER OF SEQ ID NOS: 122

/ SOFTWARE: PatSeq for Windows Version 4.0

/ SEQ ID NO 16

/ LENGTH: 396

/ TYPE: PRT

/ ORGANISM: Homo sapien

US-09-802-640-16

Query Match 4.2%; Score 93.5; DB 11; Length 396;

Best Local Similarity 20.1%; Pred. No. 3.7;

Matches 80; Conservative 63; Mismatches 155; Indels 101; Gaps 18;

QY 38 VITPKGSLKESRIKISSPOFANGWEY--HISNTE-----KPEVQSRINPLEATI 89

Db 7 VLTALVAVAGARAEVSADQVATVMWDYFSQLSNNKAEVHLQKSELTOQLNALFQDKL 66

QY 90 FIVLAIQPTAEADLE--IIYSDPGYHSQEDTETKTSNGEKTFLYHSRAITEVEKTGL 147

Db 67 GEVNTY-----AGDLQKLVPPA-----TE-----LHERLADSEK--L 98

QY 148 GSSAGL-VSVVATSLSHFIPNVISTN-KDILHNAQIAHCYA-QKKIGSGFDVATAIYG 204

Db 99 KEIGKELEELRRLPH--ANEVSQKIGDNLRELQORLEPYADQLRQVNTQAEQLRQ 156

QY 205 LIYRRFPQPALINDVQVLESDEPEKPEPTELKLIENMEKHEKRCPLPYGIKLMGDVYG 264

Db 157 LTPYAQRMERVLRENADSLQASLRPHADELAKKIDQVVELKGRLT-PYA----- 205

QY 265 GSETPKLVSRLQWKKEKEPESSVYVDQLSANLQFMKELREMERREKYSDPEPTYIKELDH 324

Db 206 -----DEFKVKIDQT-----VEELRSLAPYADQ-----EKLNH 236

QY 325 SVEPLTVAIKNIRKGLQALTQKSEVPPIEPDVOTQLDRQCEIPGCVGVPG-----A 377

Db 237 OLEGLTFQMKNAEELKARISAS---AEELRQRLAPLAEDVGNLKGNTGELQKSLAEL 292

QY 378 GGYDAIAVLVLNVOVGNFKQKTLNPDYFHNYYWVLEE 416

Db 293 GGH-----LDQVBEFFRRRVEPYGENFKALVQOMEQ 324

Search completed: August 8, 2003, 09:48:53

Job time : 56 secs